UNIVERSITÀ DI PISA

Scuola di Dottorato in Ingegneria "Leonardo da Vinci"

Corso di Dottorato di Ricerca in
SICUREZZA NUCLEARE E INDUSTRIALE

Tesi di Dottorato di Ricerca

APPLICAZIONE DI TECNICHE ALTERNATIVE PER LA PREVISIONE DELL'INQUINAMENTO DELL'ARIA

Previsione dell'inquinamento da CO dovuto al traffico urbano

Autore:

Alessia Babboni _____

Relatori:

Prof. Marco Nicola Mario Carcassi _____

Ing. Giancarlo Fruttuoso _____

Anno 2007

INDICE

PARTE II: SVILUPPO DEL MODELLO COMPOSTO, INDIVIDUAZIONE DELLA METODOLOGIA, CONCLUSIONI

SOMMARIO

*Il lavoro svolto costituisce un esempio di sviluppo di una metodologia da applicare su scala locale, di tipo '*user friendly*, tesa a valutare l'inquinamento atmosferico in aree urbane dove la sorgente "traffico veicolare" costituisce una delle maggiori cause d'inquinamento.*
L'esigenza di sviluppare una metodologia applicabile alle diverse realtà locali per effettuare studi di inquinamento dell'aria ha portato ad adottare un approccio di tipo modellistico innovativo, in particolare per quei casi in cui le caratteristiche meteorologiche delle città esaminate (es. brezze marine, effetti canyon, etc), associate ad un profilo urbanistico complesso, non rendano possibile l'uso dei codici gaussiani normalmente usati in tale ambito. La scelta è ricaduta sull'uso di tecniche ibride neuro-fuzzy, in grado di trattare ambiti in cui i modelli classici porterebbero ad una attività eccessivamente onerosa per la risoluzione del problema.
Lo strumento a cui è stato fatto riferimento per l'applicazione della logica ibrida è il codice fuzzy-Tech 5.54e, della fuzzyTECH® Inform GmbH che utilizza un modello del tipo Takagi-Sugeno.
Tra i diversi inquinanti emessi dai veicoli l'attenzione è stata rivolta al monossido di carbonio, in quanto la sua relativamente lunga permanenza nei bassi strati dell'atmosfera lo rende un ottimo indice della qualità dell'aria.
La ricerca svolta può essere divisa in due parti principali:

Parte I

Dopo un'analisi delle teorie fuzzy, neurale e dei sistemi ibridi per acquisire le nozioni base necessarie per l'esecuzione della ricerca, l'attenzione è stata rivolta alla:

- *costruzione di un caso ideale per testare la validità dei risultati del codice fuzzyTECH 5.54e operando nelle stesse condizioni di un codice gaussiano precedentemente validato;*
- *analisi di sensibilità del modello.*

Parte II

Individuato un possibile sito in cui svolgere l'analisi e dopo un esame fisico dello stesso (analisi morfologica e climatica, uso del suolo e caratterizzazione della sorgente di inquinamento da traffico) per determinare le variabili maggiormente significative per la definizione del contesto, lo studio ha riguardato:

- *l'analisi dei dati acquisiti, la validazione del modello e l'analisi degli errori ottenuti;*
- *la riduzione del set di dati iniziali in sub-sets utilizzando criteri di ottimizzazione emersi durante la fase di assessment;*
- *la messa a punto di sottomodelli che impiegano i subsets di cui al punto precedente;*
- *l'utilizzo della teoria del MultiDimensional Scaling per la localizzazione di nuovi dati e per la verifica dell'attendibilità dei risultati ottenuti durante la fase prognostica;*
- *la messa a punto della metodologia da applicare nel caso di analisi dell'inquinamento dell'aria utilizzando un modello "neuro-fuzzy".*

L'analisi così svolta ha portato a definire una metodica di approccio al problema, che, seppur svolta in un contesto specifico, non perde di validità generale.
Tale metodologia può essere suddivisa in due parti fondamentali:

- *una prima riguardante un approccio mirato alla valutazione dello stato dell'ambiente al momento dell'analisi;*
- *una seconda rivolta alla previsione dell'inquinamento nel sito esaminato.*

Rientrano nel primo punto tutta una serie di indicazioni preliminari, la cui applicazione assicura l'individuazione di set di dati significativi, privi di ridondanze e "outliers", nonché di variabili necessarie per il contesto analizzato. Particolare attenzione è rivolta all'individuazione delle Membership Functions e, nel caso di utilizzo, alla messa a punto del codice fuzzyTECH5.54e
Rientrano invece nel secondo punto tutta una serie di procedure da seguire onde poter individuare il sottomodello di riferimento nella fase prognostica.
L'obiettivo principale della ricerca è stato pertanto la messa a punto e la convalida di uno strumento, di semplice utilizzo, in grado di coadiuvare le autorità locali nell'applicazione della normativa di tutela della qualità dell'aria ambiente in aree urbane.

ABSTRACT

The present study offers an example on how an easy-to-use methodology to be applied at the local level can be developed to estimate air pollution in urban areas where motor-vehicle emissions represent one of the major causes of pollution.
The need to develop a methodology to be employed into different local areas to perform air pollution studies has led to the adoption of an innovative modelling approach in particular when weather conditions of cities (i.e. sea breezes, canyon effect etc) associated with complex urban profiles make it impossible to use Gaussian codes.
Hybrid “neuro-fuzzy” techniques that can be properly used in situations, in which the classical models would be too costly for the solution of the problem, were selected.
Fuzzy-Tech 5.54e, created by fuzzyTECH® Inform GmbH from a Takagi-Sugeno model, is the hybrid logic software that was used.
Carbon monoxide was selected among the different road traffic pollutants because it remains for a relatively long period of time in the lower atmosphere layers making it an excellent air quality index.
Study carried out can be split into two main parts:
Part I
After analysis of neural and fuzzy theories and of hybrid systems to acquire further knowledge for the purposes of the present research, attention was focused on:
-building an ideal case to test validity of the fuzzyTECH 5.54e results obtained by operating under the same conditions of a Gaussian code previously validated;
-analysing sensitivity of the model.
Part II
After a suitable site to carry out analysis was identified and its physical properties were investigated (morphology and climate conditions, use of soil and characterization of road traffic pollution sources) to determine the most important variables for defining the context, the study concentrated on:
- data analysis, model validation and error analysis;
- reducing initial data set into sub-sets using optimization criteria emerged during the assessment phase;
- creating sub-models using the above sub-sets;
- using the MultiDimensional Scaling theory to localise new data and to verify reliability of results obtained during the prognosis phase
- creating a methodology to be applied for air pollution analysis by using the neuro-fuzzy model
Although this new approach method was conducted in a specific context it still has a general validity.
This methodology can be divided into two main fundamental parts consisting of:
- an approach evaluating the actual environment condition when performing the analysis
- a pollution forecast in the site examined.
With reference to the first part, application of preliminary indications ensures both identification of substantial data sets, free of redundancies and outliers, and of variables essential to the context being analysed. Special emphasis is placed on the identification of Membership Functions and, in case of use, on the optimization of the fuzzyTECH5.54e code.
Part two focuses on procedures to be followed to identify reference sub-models in the prognosis phase.
Main objective of the present study was adapting and validating an easy-to-use tool to help local authorities to ensure enforcement of air quality regulations in urban areas.

Considerazioni introduttive

All'inizio del lavoro si ritiene opportuno, data la complessità dell'argomento trattato, anticipare una sintetica descrizione dell'intera attività svolta, perché il lettore possa affrontare i singoli argomenti avendo già chiara la logica seguita durante lo studio.

1. Negli ultimi decenni l'inquinamento atmosferico si è rivelato un vero e proprio problema per le aree urbane italiane, soprattutto perché accanto a un sempre crescente incremento della mobilità quasi mai è stata posta l'attenzione ad un pari sviluppo delle infrastrutture di trasporto.

Le città svolgono un ruolo da protagoniste, sia come responsabili che come vittime dell'inquinamento atmosferico. La maggior parte della popolazione infatti vive e lavora in città ed è costretta ad entrare in contatto con emissioni di sostanze tossiche.

Le risorse disponibili in materia di qualità dell'aria rappresentano uno dei fattori determinanti per il controllo della capacità di ogni amministrazione municipale di attuare con successo misure volte al miglioramento della qualità dell'aria urbana. E poiché lo sviluppo sostenibile consiste nella ricerca simultanea della competitività economica, della coesione sociale e della sostenibilità ambientale, lo stesso può essere raggiunto solamente tramite un approccio integrato, in modo da far sì che diversi settori strategici vengano armonizzati e seguano un obiettivo comune.

La normativa vigente del resto (Codice dell'ambiente recentemente approvato - DLgs. 152 /2006 e s.m.i.) pone l'accento sulla necessità di operare basandosi sulle migliori tecniche disponibili (MTD) compatibili con gli strumenti di pianificazione e programmazione del territorio e l'Autorizzazione Integrata Ambientale (AIA) sembra essere uno strumento in grado di assicurare un adeguato controllo delle principali sorgenti di inquinamento, nel quadro di una corretta pianificazione urbana, anche se è da sottolineare che ad esso contribuiscono in maniera sostanziale il traffico veicolare ed, in inverno, il riscaldamento degli edifici.

L'esigenza di sviluppare una metodologia applicabile alle diverse realtà locali per effettuare studi di inquinamento dell'aria ha portato ad adottare un approccio di tipo modellistico innovativo, in particolare per quei casi in cui le caratteristiche meteorologiche delle città esaminate (es. brezze marine, effetti canyon, etc), associate ad un profilo urbanistico complesso, non rendano possibile l'uso dei codici gaussiani normalmente usati in tale ambito. La scelta è ricaduta sull'uso di tecniche ibride neuro-fuzzy, in grado di trattare ambiti in cui i modelli classici porterebbero ad una attività eccessivamente onerosa per la risoluzione del problema.

Il lavoro svolto costituisce un esempio di sviluppo di una metodologia da applicare su scala locale relativamente 'easy to use', tesa a valutare l'inquinamento atmosferico in aree urbane dove la sorgente "traffico veicolare" costituisce una delle maggiori cause d'inquinamento.

Lo strumento a cui è stato fatto riferimento per l'applicazione della logica ibrida è il codice fuzzy-Tech 5.54e, della *fuzzy*TECH® Inform GmbH che utilizza un modello del tipo Takagi-Sugeno.

Tra i diversi inquinanti emessi dai veicoli l'attenzione è stata rivolta al monossido di carbonio, in quanto la sua relativamente lunga permanenza nei bassi strati dell'atmosfera lo rende un ottimo indice della qualità dell'aria.

La ricerca svolta può essere divisa in due parti principali:

Parte I

Dopo un'analisi delle teorie fuzzy, neurale e dei sistemi ibridi per acquisire le nozioni base necessarie per l'esecuzione della ricerca (vd. App.1), l'attenzione è stata rivolta a:

- la costruzione di un caso ideale per testare la validità dei risultati del codice fuzzyTECH 5.54e operando nelle stesse condizioni di un codice gaussiano (ISC3-st);
- l'analisi di sensibilità del modello.

Parte II

Individuato un possibile sito per cui svolgere l'analisi e dopo un esame fisico dello stesso (analisi morfologica e climatica, uso del suolo e caratterizzazione della sorgente di inquinamento da traffico) per determinare le variabili maggiormente significative per la definizione del contesto, lo studio ha riguardato:

- l'analisi dei dati acquisiti, la validazione del modello e l'analisi degli errori ottenuti;
- **la** riduzione del set di dati iniziali in sub-sets utilizzando criteri di ottimizzazione emersi durante la fase di assessment;

- l'utilizzo della teoria del MultiDimensional Scaling (vd App.9) per la localizzazione di nuovi dati e per la verifica dell'attendibilità dei risultati ottenuti durante la fase prognostica;
- la messa a punto della metodologia da applicare nel caso di analisi dell'inquinamento dell'aria utilizzando un modello "neuro-fuzzy".

2. In particolare nella **prima parte** è stata reperita ed analizzata una congrua documentazione sia di base che relativa a casi concreti dell'applicazione delle metodiche sopra dette, che hanno evidenziato come la logica booleana mostri spesso evidenti limiti di applicabilità alla realtà di tutti i giorni.

La logica fuzzy risulta un ottimo strumento di gestione della polivalenza e della vaghezza, ammettendo una struttura formale che ne permette una successiva rappresentazione numerica (vd All.1). Inoltre presenta tre caratteristiche principali:
1) l'utilizzo di variabili linguistiche i cui valori non sono numeri, ma parole in linguaggio naturale o artificiale, al posto o in aggiunta alle variabili numeriche;
2) l'individuazione di semplici relazioni tra variabili attraverso affermazioni condizionali fuzzy;
3) la caratterizzazione di relazioni complesse tra variabili attraverso algoritmi fuzzy.

La generalizzazione dalla logica booleana a quella fuzzy passa attraverso l'estensione del concetto di appartenenza di un elemento ad un insieme .

Nella prima un oggetto appartiene o non appartiene ad un gruppo e la sua caratterizzazione può essere fatta mediante un semplice elenco dei suoi componenti.

Nella logica fuzzy un elemento fa parte di un insieme con un certo grado di "somiglianza"; tale grado può essere espresso da una funzione (membership function, MF) i cui valori sono compresi tra un minimo ed un massimo che risultano rispettivamente 0 e 1 nel caso in cui tale funzione è normalizzata.

Un sistema che opera utilizzando la logica fuzzy esegue tre passi fondamentali:

1. la "fuzzificazione degli inputs
2. l'inferenza fuzzy
3. la defuzzificazione degli outputs

Il procedimento attraverso il quale le variabili di ingresso (es. temperatura, pressioni, ecc.) vengono convertite in misure fuzzy ("fuzzificazione") è effettuato attraverso le membership functions predefinite per quel contesto (Fig.a).

All'aumentare delle misure fuzzy e del numero di variabili aumentano i casi derivanti dalla combinazione di tutti i possibili "accostamenti linguistici", casi che individuano tutte le leggi che possono regolare il sistema. L'individuazione delle leggi maggiormente significative, necessarie per definire il comportamento del sistema, avviene attraverso l'affidamento di "pesi" tratti dalla casistica emersa dall'analisi dei dati reperiti (fase di training). In tal modo ogni regola assume una importanza relativa diversa, a seconda della sua frequenza di accadimento e risulterà tanto più attendibile quanto più grande sarà il numero di "misure" tratte dalla realtà.

Una volta ricavati i "pesi", eventuali nuovi dati saranno trattati in base al loro grado di somiglianza rispetto alla base di conoscenza che il sistema ha imparato a trattare, ricavando l'output come "interpolazione dei valori più "vicini"

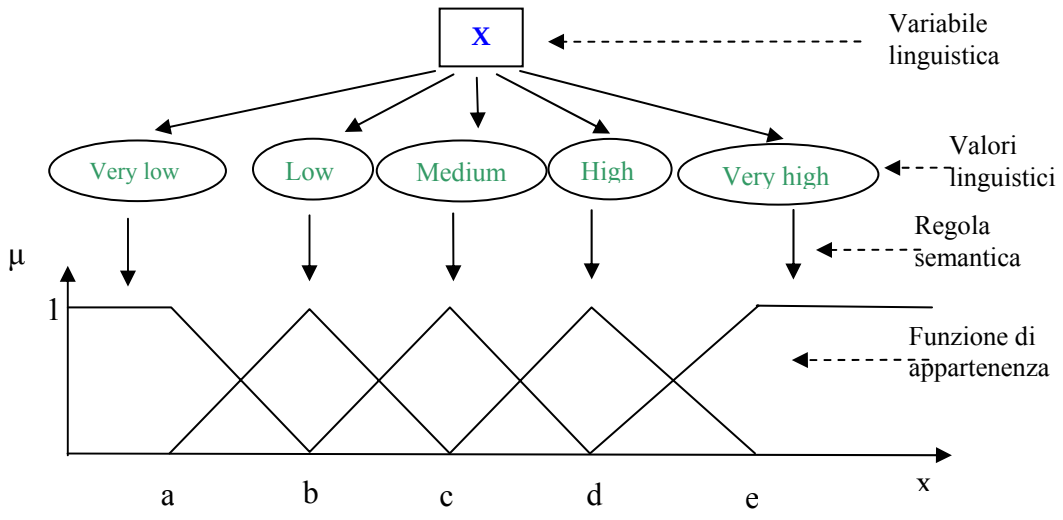

Fig. a:Relazione tra variabile linguistica e funzione di appartenenza

3. Contemporaneamente allo studio dei sistemi ibridi, è stata svolta l'analisi preliminare del codice fuzzy-TECH 5.54e, ed è proprio durante questo approccio preliminare che è emerso il suo limite sostanziale: il numero di regole automaticamente generate dal codice è inferiore a 1200; nel caso in cui le partizioni delle variabili implichino un numero di regole ≥ 1200 il codice richiede l'implementazione manuale.
In questo modo la possibilità di adottare member functions che meglio rappresentino l'andamento degli inputs e degli outputs resta vincolata inesorabilmente alla divisione in intervalli delle variabili; operare un "fitting" troppo spinto può comportare un superamento della soglia delle regole, quindi l'impossibilità pratica dell'utilizzo del codice.

Nel corso dello studio spesso è stato necessario trovare un buon compromesso tra il numero di partizioni e l'andamento delle membership functions di ogni singola variabile.

A questo punto, onde poter validare la metodologia rispetto all'approccio classico (i.e. ISC3), è emersa l'esigenza di definire due aspetti sostanziali:

a) su quali risultati sarebbe stata verificata la capacità del fuzzyTech 5.54e, cioè quali risultati di ISC3 avrebbero costituito il termine di paragone?
b) quando il codice neuro-fuzzy avrebbe potuto essere ritenuto "affidabile"?

Per quanto riguarda il punto a), qualsiasi output avrebbe avuto la stessa valenza ai fini dello studio, per cui l'attenzione è stata rivolta alla concentrazione di CO lungo un tratto stradale di una città ipotetica, con una situazione meteo creata a priori e fattori di emissione autoveicolare tratti da un precedente lavoro svolto nel Dipartimento di Costruzioni Meccaniche, Nucleari e della Produzione dell'Università di Pisa.

Il punto b) invece ha comportato una impegnativa ricerca in letteratura, per cercare dei criteri di similitudine rispetto alle metodiche "usuali" che fornissero almeno l'ordine di grandezza dell'errore ritenuto accettabile. In questo settore non tutti gli autori sono concordi nel definire una soglia di accettabilità, dato che spesso i risultati dipendono dal tipo di argomento trattato (controllo industriale, analisi ambientale, valutazioni epidemiologiche, etc.) e dal tipo di rete neuro-fuzzy impiegata. Tuttavia si è ritenuto accettabile, in prima istanza, un errore tra i dati calcolati e quelli misurati dell'ordine del 20%.

4. Per creare il caso ideale ipotizzato in precedenza sono state assunte alcune ipotesi iniziali:

- che potessero verificarsi tutte le classi di stabilità da A a F con uguale probabilità (unitaria) condizionata solo dalla velocità del vento, che si è ipotizzato variare come mostrato dalla tabella 1

Tab.1: Significatività di esistenza di una classe di stabilità in funzione della velocità del vento**.**

	v≤2	2<v≤3	3<v≤4	4<v≤5	5<v≤6	v>6
A	1	1	0	0	0	0
B	1	1	1	0	0	0
C	0	1	1	1	1	1
D	0	0	1	1	1	1
E	0	1	1	0	0	0
F	1	1	0	0	0	0

- che ad ogni caso potesse essere associata una direzione del vento variabile tra 0 e 360° (andamento a gradino con step di 22,5°).

In totale si sono trattati 288 casi possibili.

In base a tali ipotesi sono stati generati mediante ISC3 i dati di concentrazione di CO, una parte dei quali è stata impiegata per la costruzione della rete "neuro-fuzzy"ed una parte dei quali è stata utilizzata per la sua validazione.

L'uso di ISC3, come modello di riferimento per la generazione del set di dati occorrenti per il training e la validation della rete neuro-fuzzy, ha comportato nello stesso tempo l'individuazione dei possibili inputs del codice fuzzyTECH 5.54e.

Si è ritenuto che le due metodologie potessero essere considerate equivalenti da un punto di vista dell'affidabilità solo se, utilizzando gli stessi inputs o una parte di questi, si fossero ottenuti gli stessi risultati entro il margine di errore accettabile (20%). Ciò ha comportato l'esecuzione di numerose prove nelle quali sono stati fatti variare tutti i possibili parametri del codice (numero di inputs, metodi di defuzzificazione, metodi di attribuzione dei pesi alle regole, numero di allenamenti, etc.).

Per effettuare l'allenamento della rete neuro-fuzzy in modo significativo e veloce è possibile operare in tre modi (vd. Cap.1), ossia attraverso l'ottimizzazione:

- del metodo di allenamento (Learn Methods);
- delle caratteristiche di allenamento (Learn Parameters);
- delle condizioni di fine allenamento (Stop Conditions)

Tra i metodi di allenamento il codice fuzzyTECH5.54e distingue i sistemi standard e quelli "batch"; i metodi standard prevedono l'ottimizzazione dei parametri per un campione dell'intera popolazione; i metodi "batch" invece calcolano sia gli errori che i gradienti su tutti i campioni ed operano l'aggiornamento dei parametri solo dopo che ogni iterazione viene conclusa.

Ai primi viene generalmente attribuito un miglior comportamento nei risultati, mentre ai secondi un miglior comportamento nell'individuazione del trend; la scelta per lo specifico contesto, avendo come obiettivo la quantificazione dell'errore, è stata rivolta ai metodi standard e tra questi a quello "Random" (vd.Cap. II).

Con la dizione "caratteristiche di allenamento" ci si riferisce ai mezzi con cui far "emergere" le regole più importanti; essi riguardano:

- il numero di neuroni "vincenti" ("winner neurons"), depositari delle regole fuzzy maggiormente significative;
- il valore addizionato al peso delle regole dei neuroni vincenti e sottratto alle regole di quelli perdenti (Step Width –DoS);
- il valore da incrementare ai vari dati, espresso in percentuale, per analizzare il nuovo set di dati appartenenti al campione esaminato (Step Width –Term)

Volendo utilizzare un approccio simile al "competitive learning" di una rete puramente neurale, in quanto comporta l'uso di algoritmi di calcolo più semplici, quindi minor tempo impiegato, il numero di neuroni vincenti è stato settato al valore unitario. In questo modo solo una regola per volta viene modificata. I valori di Step With sono stati invece fatti variare in un range che va dal 50 al 1%.

I migliori risultati si sono avuti utilizzando:

- una parte degli inputs necessari per ISC3;
- un metodo random per l'estrazione dei dati impiegati nella fase di training;
- un metodo random per l'affidamento dei pesi iniziali alle regole fuzzy;
- un allenamento dinamico con soglia di massimo errore 30% e soglia di errore medio 0,1%.;
- Step with Dos , Step with Term e Winner pari a 0,1, 1 e 1 rispettivamente.

In tutti i casi la "stop condition" (condizione di fine allenamento) è sempre stata ottenuta per il conseguimento del numero massimo di iterazioni preimpostato. L'errore massimo calcolato tra i dati prodotti dal fuzzyTECH 5.54e e quelli prodotti da ISC3 è rimasto quasi sempre entro un margine del 15%, alcune volte tra il 15% e il 20%.

5. Nella **seconda parte** è stata svolta, come detto in precedenza, l'applicazione ad un caso concreto. La difficoltà incontrata è stata quella di muoversi in uno spazio multi-dimensionale; ogni osservazione infatti è caratterizzata da 6 valori (5 input, 1 output) che possono essere trattati alla stregua di una matrice; riconoscere la posizione reciproca di questi non è né semplice né immediato.

Del resto, per ottenere risultati attendibili dal codice, il training e la validation devono avvenire con dati che appartengono ad una stessa regione di spazio poiché la logica neurale interpola "bene", ma estrapola "male". Si è pensato di iniziare semplicemente plottando tutte le variabili, analizzandone il comportamento ed eventualmente confrontandone l'andamento nei due mesi disponibili (giugno e luglio 2002). Ipotizzando l'assenza di grandi errori di misura, essendo del tutto paragonabili i mesi considerati da un punto di vista meteo e di traffico urbano, gli andamenti avrebbero dovuto essere similari, e così è stato. Plottando poi le variabili a coppie, i punti ottenuti avrebbero dovuto occupare uno spazio ben definito su ogni grafico, in entrambi i mesi considerati; in più se un mese avesse sempre contenuto, per ogni coppia plottata, la rispettiva dell'altro mese, avremmo potuto dire che i dati del primo "contenevano spazialmente" i dati del secondo. Facendo poi una verifica di eventuali "coincidenze" e la loro successiva eliminazione, sarebbe stato possibile dividere i dati in due set significativi. Purtroppo però la "delimitazione spaziale" di un mese rispetto all'altro non si è verificata, anzi alcune volte l'andamento di una coppia di variabili del mese di giugno conteneva l'andamento della rispettiva coppia del mese di luglio; in altre occasioni accadeva il contrario. Una simile analisi non poteva quindi fornire alcuna informazione circa la creazione di due set di dati significativi. Parallelamente ad una indagine di tipo cluster è stato effettuato un tentativo con il codice fuzzy-TECH 5.54e mescolando tutti i dati disponibili, estraendone i primi mille per l'allenamento e i rimanenti per la validazione.

Utilizzando le condizioni migliori ottenute nella messa a punto del modello e confermate durante l'assessment, l'errore medio e l'errore massimo commessi dal codice nel processare i mille dati scelti per il training rimanevano al di sotto del 10%. Tuttavia nella fase di validazione tali errori aumentavano fino a valori 4-5 volte maggiori.

E' stato ipotizzato che nel primo caso la rete avesse "imparato troppo" e si è proceduto con un secondo tentativo diminuendo il numero di iterazioni eseguite dal codice nella fase di training. Il risultato però

ha visto aumentare notevolmente gli errori nella fase di training, senza intervenire minimamente sugli errori della fase di validazione.

6. Uno studio mirato ha portato alla formulazione di una nuova impostazione del problema, con l'obiettivo di ridurne la complessità senza tuttavia perdere informazioni rilevanti ai fini del contesto. Ciò poteva essere conseguito in due modi:
1) *indiretto,* attraverso l'uso di
a) membership functions più complesse (es. una maggiore ripartizione lungo l'intervallo di variazione per incrementare il grado di precisione);
b) una metodica di "search for structure in data" (limitazione delle regole fuzzy mediante una analisi approfondita dei dati);
2) *diretto* con l'uso di sottomodelli finalizzati

Per quanto riguarda il punto a) il limite di ammissibilità delle regole fuzzy posto dal codice (<1200) ha impedito una qualsiasi ulteriore ripartizione a meno dell'esclusione di una variabile.

Per quanto riguarda il punto b), un'analisi approfondita dei dati ha portato all'individuazione di regole "più importanti", ma i risultati ottenuti implementando nel codice solo queste ultime regole sono stati non significativi.

L'attenzione è stata perciò rivolta all'uso di sottomodelli, ossia modelli che simulano realtà circoscritte dove la variazione dei parametri rimane contenuta. Ma quanto contenuta?

In prima istanza si è pensato di suddividere tutti i dati in quattro gruppi, in base alla direzione del vento: 0°÷90°, 90°÷180°,180°÷270°, 270°÷360°. Tale ripartizione continuava a non fornire risultati migliori. E' risultato perciò necessario riesaminare il tutto partendo dalla ipotesi che i dati ricoprissero regioni di spazio ben definite, limitate ma non contigue. In queste gli intervalli di variazione delle variabili dovevano essere circoscritti in modo da consentire una maggiore precisione nella loro trattazione. Poiché durante la fase di validazione è stato rilevato che risultati maggiormente significativi venivano ottenuti quando, per ogni variabile, il rapporto tra la deviazione standard e la media rimaneva inferiore al 40% (vd Cap.II), si è iniziato nell'imporre tale condizione al parametro più indicativo per la determinazione del monossido di carbonio, ossia il flusso veicolare..
In particolare per definire i sottogruppi sono stati presi in esame due fattori:

- la numerosità dei dati
- la deviazione standard del flusso veicolare

Entrambi i parametri si è dimostrato che giocano un ruolo fondamentale per la convergenza degli indici di qualità del codice. Il codice infatti ha mostrato di raggiungere meglio i target (deviazione media <2%; deviazione massima <5%, definiti nel primo anno durante la validazione della metodologia) se:

- la numerosità delle osservazioni rimaneva inferiore a 50, qualora la deviazione standard del flusso veicolare rispetto alla media era ≤ 30%;
- la numerosità delle osservazioni aumentava a circa 60, se la deviazione standard del flusso veicolare rispetto alla media era > 30%;

La fase di training, effettuata con l'85% delle osservazioni per ogni sottogruppo, si è conclusa molto spesso per il conseguimento della deviazione media <0,1% e della deviazione massima <1%, altre volte per il raggiungimento del numero massimo di iterazioni pre-impostate. In ogni caso è risultato sempre un errore medio <2% ed un errore massimo < 5% . Utilizzando il 15% rimanente di ogni sottogruppo, la fase di validation ha mostrato che l'errore medio e massimo sono rimasti, nel 95% dei casi, al di sotto dell'8%. Per il 5 % rimanente, per i quali gli errori raggiungono anche il 120%, è stato verificato che i dati di input appartenevano a porzioni di spazio dove direzione del vento e flusso veicolare assumevano i valori limiti dei propri intervalli di variazione.

Ne è derivato così un modello composito a 56 sottomodelli, la cui scelta nella successiva fase prognostica è dettata dai valori della direzione del vento e del flusso veicolare presi in esame.

7. Volendo utilizzare il modello in fase prognostica, il problema emerso a questo punto è stato quello di stabilire se il vettore di input a disposizione costituisce o no un "border line" e come tale se i risultati ottenuti possono essere ritenuti affidabili o meno.

Per valutare la bontà dei risultati ottenuti durante la fase prognostica, si è pensato di utilizzare la teoria del *MultiDimensional Scaling,* secondo la quale ogni elemento a n dimensioni può essere rappresentato come un punto in un adeguato spazio metrico a p dimensioni (con p<<n) tale che la nuova configurazione di punti rispetti nel miglior modo le distanze originarie (vd All. 9). In questo modo la rappresentazione grafica dei dati fornisce un utile strumento per l'analisi dell'appartenenza di nuovi dati alla regione di spazio individuata dal subset utilizzato per la messa a punto del sottomodello, quindi per la sua proficua utilizzazione.
L'analisi così svolta ha portato a definire una metodica di approccio al problema, che, seppur svolta in un contesto specifico, non perde di validità generale.

Tale metodologia può essere suddivisa in due parti fondamentali:

- una prima riguardante un approccio mirato alla valutazione dello stato dell'ambiente al momento dell'analisi;
- una seconda rivolta alla previsione dell'inquinamento nel sito esaminato.

Rientrano nel primo punto tutta una serie di indicazioni preliminari, la cui applicazione assicura l'individuazione di set di dati significativi, privi di ridondanze e "outliers", nonché di variabili necessarie per il contesto analizzato. Particolare attenzione è rivolta all'individuazione delle Membership Functions e, nel caso di utilizzo, alla messa a punto del codice fuzzyTECH5.54e

Rientrano invece nel secondo punto tutta una serie di procedure da seguire onde poter individuare il sottomodello di riferimento nella fase prognostica.

I. IL CODICE fuzzy-TECH 5.54e

1.1 Premessa

Le risorse disponibili in materia di qualità dell'aria sono elementi particolarmente importanti, infatti rappresentano uno dei fattori determinanti per il controllo della capacità di ogni amministrazione municipale di attuare con successo misure volte al miglioramento della qualità dell'aria urbana. E poiché lo sviluppo sostenibile consiste nella ricerca simultanea della competitività economica, della coesione sociale e della sostenibilità ambientale, lo stesso può essere raggiunto solamente tramite un approccio integrato, in modo da far sì che diversi settori strategici vengano armonizzati e seguano un obiettivo comune.
Del resto la normativa vigente (Codice dell'ambiente recentemente approvato -DLgs. 152- /2006 e s.m.i.) pone l'accento sulla necessità di operare basandosi sulle migliori tecniche disponibili (MTD) compatibili con gli strumenti di pianificazione e programmazione del territorio e l'Autorizzazione Integrata Ambientale (AIA) sembra essere uno strumento in grado di assicurare un adeguato controllo delle principali sorgenti di inquinamento, nel quadro di una corretta pianificazione urbana, anche se è da sottolineare che ad esso contribuiscono in maniera sostanziale il traffico veicolare ed, in inverno, il riscaldamento degli edifici.
L'esigenza di sviluppare una metodologia applicabile alle diverse realtà locali per effettuare studi di inquinamento dell'aria ha portato ad adottare un approccio di tipo modellistico.
In particolare l'interesse è stato rivolto a quei casi in cui le caratteristiche meteorologiche delle città esaminate (es. brezze marine, effetti canyon etc), associate ad un profilo urbanistico complesso non rendano possibile l'uso di codici gaussiani. La scelta è ricaduta perciò sull'uso di tecniche ibride, in grado di trattare ambiti in cui i modelli classici porterebbero ad una attività eccessivamente onerosa per la risoluzione del problema.
Il lavoro svolto costituisce un esempio di sviluppo di una metodologia da applicare su scala locale, 'easy to use', tesa a valutare l'inquinamento atmosferico in aree urbane dove la sorgente "traffico veicolare" costituisce una delle maggiori cause d'inquinamento.
Lo strumento a cui è stato fatto riferimento per l'applicazione della logica ibrida è il codice fuzzy-Tech 5.54e, della *fuzzy*TECH® Inform GmbH che utilizza un modello del tipo Takagi-Sugeno.

1.2 Fuzzy associative memories (FAMs)

Il modello sopra detto ha un'architettura che sfrutta le "fuzzy associative memories" /1/ /2/ /3/ ed è costituito da cinque blocchi funzionali (vd.Fig.1.1):

- una "rule base" contenente le regole fuzzy del tipo "if…then";
- un "database" che definisce le funzioni di appartenenza dei fuzzy sets usati nelle regole fuzzy;
- un blocco di decisione che attua le operazioni sulle regole;
- un'interfaccia di fuzzificazione che trasforma i valori crisp di input in termini linguistici con grado di appartenenza al set esaminato;
- un'interfaccia di defuzzificazione che trasforma i risultati fuzzy in valori crisp.

I primi due blocchi sono generalmente noti come "base della conoscenza"

(Knowledge base).

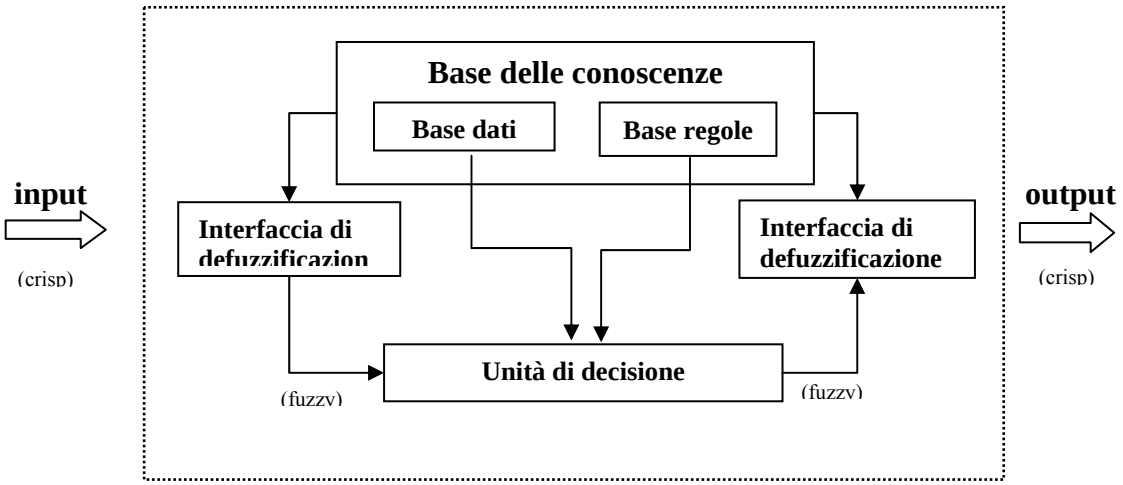

Fig. 1.1: Fuzzy Inference System /1/

Ogni regola fuzzy è considerata costituita da due parti: la parte *antecedente* caratterizzata dall'operatore "IF" e la parte *conseguente* caratterizzata da "THEN" /1/.

Le operazioni svolte dal sistema possono essere suddivise in:

- fuzzificazione: le variabili di input sono comparate alle funzioni di appartenenza per ottenere misure di compatibilità per ogni label linguistica;
- combinazione: attraverso specifici operatori, generalmente di moltiplicazione o di minimo, assegna un peso alla parte antecedente della regola
- calcolo: genera la parte conseguente di ogni regola in funzione dei pesi della parte precedente
- aggregazione: associa i risultati ottenuti nella fase di "calcolo" e li trasforma in valori crisp

Durante la fase di "combinazione" avviene quindi l'affidamento di un peso alla parte precedente di ogni regola. Per semplicità possiamo immaginare di accumulare tutti i valori in una matrice, che sarà chiamata "matrice di connessione"; quando la rete neuro-fuzzy viene allenata, ogni qualvolta un'informazione richiama una qualsiasi regola non fa altro che andare a modificare il peso relativo all'interno della matrice. Da tale "associazione" deriva l'acronimo FAMs (fuzzy associative memories), volto a sottolineare l'intima relazione tra regole e pesi.
In letteratura esistono molti modelli di Fuzzy Associative Memories: spesso il meccanismo di apprendimento della rete neurale è usato per determinare non solo i pesi ma le stesse regole fuzzy, i parametri degli insiemi fuzzy, etc. Si tratta di modelli complessi che esulano dall'ambito del lavoro svolto in questa tesi di dottorato e si rimanda perciò a testi specifici.

1.3 Parametri di allenamento del codice fuzzyTECH5.54e

Per effettuare l'allenamento della rete neuro-fuzzy del codice fuzzyTECH5.54e, in modo significativo e veloce, come mostra la fig.1.2, l'utente può interagire in tre modi, ossia attraverso l'ottimizzazione:

- del metodo di allenamento (Learn Methods);
- delle caratteristiche di allenamento (Learn Parameters);
- delle condizioni di fine allenamento (Stop Conditions)

Tra i metodi di allenamento il codice fuzzyTECH5.54e distingue tra i sistemi standard e quelli "batch"; i metodi standard prevedono l'ottimizzazione dei parametri per un campione dell'intera popolazione; i metodi "batch" invece calcolano sia gli errori che i gradienti su tutti i campioni ed operano l'update dei parametri solo dopo che ogni iterazione viene conclusa.
Ai primi viene attribuito un miglior comportamento nei risultati, mentre ai secondi un miglior comportamento nell'individuazione del trend; la scelta per lo specifico contesto, avendo come obiettivo la quantificazione dell'errore, è stata rivolta ai metodi standard e tra questi a quello "Random" (vd.Cap.II).
Con la dizione "caratteristiche di allenamento" ci si riferisce ai mezzi con cui far "emergere" le regole più importanti; essi riguardano:

- il numero di neuroni "vincenti" ("winner neurons");
- il valore addizionato al peso delle regole dei neuroni vincenti e sottratto alle regole di quelli perdenti (Step Width –DoS);
- il valore da incrementare ai vari dati, espresso in percentuale, per analizzare il nuovo set di dati appartenenti al campione esaminato (Step Width –Term)

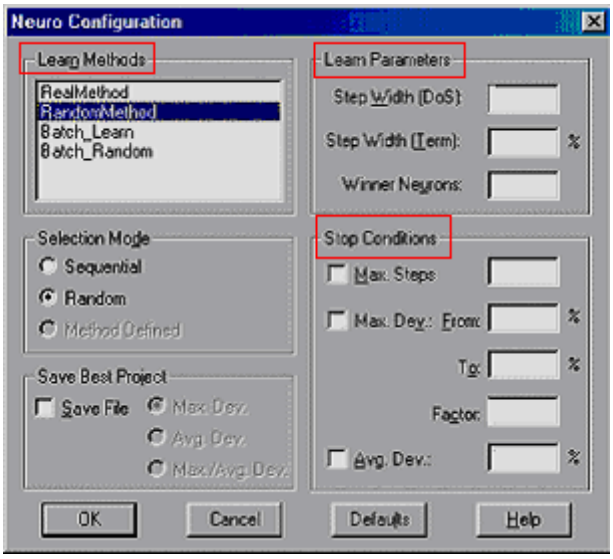

Fig.**1.**2: Parametri di allenamento del fuzzyTECH5.54e

Volendo utilizzare un approccio simile al “competitive learning” di una rete puramente neurale, in quanto comporta l’uso di algoritmi di calcolo più semplici, quindi minor tempo impiegato, il numero di neuroni vincenti è stato settato al valore unitario. In questo modo solo una regola per volta viene modificata. I valori di Step With sono stati invece fatti variare in un range che va dal 50 al 1%.
Per quanto riguarda le condizioni di fine allenamento, il codice dà la possibilità di agire sia impostando un numero di iterazioni massimo, sia fissando dei limiti riguardo l’errore medio ([*Avg Dev*]) o massimo ([*Max Dev*]) ricavato dal confronto tra i dati veri e quelli calcolati. Per diminuire i tempi di calcolo il codice può utilizzare anche un criterio dinamico per l’errore massimo (Dynamic Stop Condition): in esso viene impostato un valore abbastanza elevato ([*Max Dev from*]) ed ogni qualvolta i campioni risultano inferiori ad esso, il limite viene abbassato di una percentuale prestabilita ([*Factor*]) fino al raggiungimento della soglia ritenuta accettabile ([*To*]).
Inizialmente il numero massimo di iterazioni sono state fissate a 1000 e l’errore medio al 20%. Successivamente poiché interessava più il valore massimo dell’errore anche di un singolo campione rispetto al valore medio dell’intero set di dati, l’attenzione è stata rivolta all’individuazione dei valori ottimali per la Dynamic Stop Condition

1.4 Metodi di defuzzificazione

Il fuzzyTECH5.54e permette di convertire le quantità fuzzy in valori crisp in due modi: attraverso l’applicazione del “Center of Maximum Method” (CoM) detto anche Best Compromise (BC) (vd All.1), o attraverso l’applicazione del “Mean of Maximum Method”(MoM) (vd.All1). Nel contesto specifico il metodo CoM si è rilevato di gran lunga il migliore ottenendo, a parità delle altre condizioni (numero di inputs, forma delle MFs, numero di ripartizioni, numero di iterazioni etc), errori massimi ed errori medi pari rispettivamente a meno della metà e a meno di un sesto di quelli ottenuti con il metodo MoM (vd.Cap.II).

II. MESSA A PUNTO DELLO STRUMENTO

2.1 Premessa

Avendo come obiettivo la messa a punto di una metodologia per analizzare l'inquinamento da monossido di carbonio dovuto al traffico autoveicolare, il primo problema emerso è stato quello di indagare la validità del modello neuro-fuzzy nell'ambito specifico.
L'idea di base per poter fare l'assessment dello strumento è stata quella di creare una "realtà" nella quale i dati potessero essere ritenuti affidabili e quindi utilizzabili per le fasi di training e di validation della rete.
I molteplici studi nel campo della diffusione degli inquinanti in contesti orograficamente semplici, svolti al dipartimento di Ingegneria Meccanica Nucleare e della Produzione dell'Università di Pisa, utilizzando il codice gaussiano Industrial Source Complex Dispersion Model (ISC3) dell'EPA /4/, hanno fatto sì da poter ritenere il modello classico sufficientemente validato e perciò adatto a generare i dati di riferimento.
Si è ritenuto che i due modelli potessero essere considerati equivalenti da un punto di vista dell'affidabilità quando, applicando alla rete neuro-fuzzy gli stessi o una parte degli inputs del modello deterministico, una volta allenata con una parte dei dati generati da ISC3, si fossero ottenuti nella fase di validation gli stessi risultati entro un margine di errore accettabile (vd §2.2.1).
Seguendo quanto detto in letteratura, circa due terzi dei dati sono stati utilizzati per la fase di training della rete neuro-fuzzy, mentre un terzo sono stati impiegati per la fase di validazione.

2.2 Costruzione della realtà di riferimento

2.2.1 Scelta della soglia di accettabilità dell'errore

Nonostante una impegnativa ricerca in letteratura non sono emersi criteri di similitudine, rispetto alle metodiche "usuali", che fornissero almeno l'ordine di grandezza dell'errore ritenuto accettabile. Non tutti gli autori infatti sono concordi nel definire una soglia di accettabilità, poiché spesso i risultati dipendono dal tipo di argomento trattato (controllo industriale, analisi ambientale, valutazioni epidemiologiche etc.) e dal tipo di rete neuro-fuzzy impiegata.
A tal proposito risulta in corso un approfondito studio svolto dalle Università di Pisa, Roma, Palermo e Milano /5/ per la definizione di criteri di accettabilità dell'errore che tengano conto della difficoltà di una sua precisa valutazione. Questo aspetto è di fondamentale importanza per la validità e credibilità dei risultati e, di conseguenza, per la robustezza delle scelte normative e delle decisioni progettuali e manageriali che su di essi si basano.
Si è ritenuto tuttavia accettabile in prima istanza un errore dell'ordine del 20% calcolato tra i dati forniti da ISC3, assunti validi, e quelli del fuzzyTECH5.54e.
Il valore 20% è emerso da considerazioni svolte sull'incertezza dei dati di ISC3, i quali, nei numerosi studi pregressi, sono risultati appartenere a questa fascia.

2.2.2 Scelta del contesto di analisi

Per effettuare il confronto tra le due diverse metodologie si è fatto riferimento ad una "realtà", come già detto al § 2.1, costruita ad oc dove fossero verificate le seguenti ipotesi:

- tutte le classi di stabilità, da A a F, verificabili con uguale probabilità (unitaria) in virtù delle condizioni di velocità del vento (Tab.2.1);

	v≤2	2<v≤3	3<v≤4	4<v≤5	5<v≤6	v>6
A	1	1	0	0	0	0
B	1	1	1	0	0	0
C	0	1	1	1	1	1
D	0	0	1	1	1	1
E	0	1	1	0	0	0
F	1	1	0	0	0	0

Tab.2.1: Significatività di esistenza di una classe in funzione della velocità del vento.

- velocità appartenenti ad uno dei seguenti gruppi
  - gruppo 1: $v \leq 2$m/s;
  - gruppo 2: 2m/s $< v \leq 3$ m/s;
  - gruppo 3: 3m/s $< v \leq 4$m/s;
  - gruppo 4: 4m/s $< v \leq 5$m/s;
  - gruppo 5: 5m/s $< v \leq 6$m/s;
  - gruppo 6: $v > 6$m/s".
- direzione del vento variabile tra 0 e 360° (andamento a gradino con step di 22,5°)
- temperatura e radiazione solare misurate dalla centralina posta sulla Rotonda presso Ardenza nel mese di luglio dell'anno 2002;
- altezze di miscelamento calcolate in un lavoro pregresso svolto al Dipartimento di Ingegneria Meccanica Nucleare e della Produzione dell'Università di Pisa /3/
- tassi di emissione autoveicolari degli inquinanti fornite dall' Inventario Regionale delle Sorgenti di Emissione in Aria Ambiente (IRSE 2002)

Un tale quadro non voleva essere una rappresentazione realistica di alcun contesto, ma solo un ambito di lavoro dove poter indagare, noti i parametri meteo e le sorgenti inquinanti, le eventuali diverse potenzialità degli strumenti a disposizione.

2.2.3 Calcolo dei dati mediante ISC3

La versione short-term dell'ISC3 necessita, oltre al file di input principale in cui sono riportati i dati generali di calcolo ed i parametri emissivi delle varie sorgenti, di un altro file in cui sono descritte ora per ora le grandezze meteorologiche usate nel calcolo, tra le quali la classe di stabilità e l'altezza di miscelamento.
Generalmente l'altezza dello strato di mescolamento costituisce un elemento molto importante, dato che permette di quantificare le dimensioni della porzione di atmosfera in cui sono presenti i moti convettivi che sono significativi per il rimescolamento delle masse d'aria: Il comportamento di un effluente e le eventuali ricadute al suolo di sostanze inquinanti variano fortemente a seconda che l'atmosfera sia in equilibrio instabile, neutro o stabile. Inoltre la presenza di inversione termica può modificare sostanzialmente l'abbattimento al suolo degli inquinanti, a seconda che gli effluenti siano emessi sopra o sotto la quota di inversione /7/. Tuttavia tali considerazioni rispecchiano la realtà solo se l'inquinante viene emesso a una certa quota, classico il caso del camino di un impianto industriale. Nello specifico contesto le emissioni avvengono a livello del suolo e la misura di queste sono fatte in prossimità della sorgente. Per queste motivazioni si è ritenuto che l'altezza di mescolamento non costituisse un parametro vincolante ai fini del calcolo e non disponendo dei profili termici di radiosondaggi da cui poter risalire al valore esatto, è stato fatto riferimento all'altezza di miscelamento calcolata nel RL 1016 (03) /6/, analisi svolta prendendo a riferimento lo stesso periodo temporale, quindi con identiche caratteristiche meteo ad esclusione della direzione e della velocità del vento variabili secondo quanto detto al §2.2.1.
Per quanto riguarda la sorgente inquinante, vista la microscala a cui si è operato, è stata considerata un'intera arteria stradale, nello specifico Viale Carducci della città di Livorno. Tale scelta è stata eseguita in virtù del fatto che il codice ISC3 era stato più volte precedentemente impiegato in questo contesto per cui disponeva già di tutte le caratteristiche dell'area.
In questo modo sono stati generati mediante ISC3 173 dati di concentrazione di CO (All.2), che sono stati considerati nelle fasi successive dello studio sufficientemente affidabili tali da costituire la "realtà" di riferimento (vd §2.1)

2.2.4 Individuazione degli inputs del fuzzyTECH 5.54e

Per quanto detto al § 2.1, l'utilizzo del codice ISC3 ha permesso di identificare i possibili inputs del fuzzyTECH 5.54e ossia:

1. sorgente inquinante
2. velocità del vento
3. direzione del vento
4. temperatura
5. classe di stabilità
6. umidità
7. radiazione solare
8. altezza di miscelamento

L'uso di tutte le variabili oltre ad appesantire enormemente il programma, avrebbe richiesto un notevole impiego di risorse umane per implementare manualmente tutte le possibile regole che risultavano dalle eventuali definizioni delle membership functions. Del resto non disponendo di esperti del settore per la definizione di queste ultime sarebbe stato necessario inserire come variabile anche la loro forma.

Poiché un simile approccio non è apparso ottimale, si è proceduto sia mediante una analisi fisico-chimica che parametrica, onde risalire alla definizione delle variabili maggiormente significative.

Dalla prima è emerso che alcune di esse sarebbero state ridondanti in quanto fisicamente correlate come la temperatura e la radiazione solare, altre non significative per la diffusione a livello del suolo, come l'altezza di miscelamento (vd §2.2.3)

Si è giunti quindi alla conclusione che delle otto variabili di partenza solo la metà avrebbe dovuto essere impiegata ossia:

1. sorgente inquinante
2. velocità del vento
3. direzione del vento
4. classe di stabilità

La sorgente inquinante tuttavia doveva essere relazionata alla lunghezza dell'arteria stradale, poiché le quantità fornite dall'IRSE sono espresse come concentrazioni per Km di strada.

Questo ha comportato per l'utilizzo del codice fuzzyTECH5.54e la rielaborazione dei dati forniti dall'IRSE i cui risultati vengono evidenziati nella tab. 2.2

0re	Tasso emiss. V.Card Ovest $\mu g/m^3Km^{(*)}$	Tasso emiss. V.Card Est $\mu g/m^3Km^{(*)}$	sorgente inq totale $\mu g/m^3$
1	0,017	0,021	41,04
2	0,008	0,008	19,008
3	0,003	0,004	7,56
4	0,002	0,002	4,32
5	0,003	0,003	6,48
6	0,006	0,007	14,04
7	0,0015	0,013	30,24
8	0,045	0,049	101,52
9	0,070	0,081	163,08
10	0,049	0,064	122,04
11	0,054	0,051	113,4
12	0,059	0,057	125,28
13	0,078	0,066	155,52
14	0,070	0,059	139,32
15	0,049	0,062	119,88
16	0,051	0,056	115,56
17	0,054	0,061	124,2
18	0,066	0,066	142,56
19	0,080	0,067	158,76
20	0,088	0,068	168,48
21	0,052	0,052	112,32
22	0,027	0,028	59,04
23	0,027	0,029	60,48
24	0,027	0,026	57,24

(*)fonte dati: IRSE2002

Tab.2.2: Emissioni medie orarie di CO per Viale Carducci durante 24 ore

2.3 Studio dei parametri del codice fuzzyTECH 5.54e e analisi della loro influenza

2.3.1 Forma delle funzioni di appartenenza

Una volta definiti gli inputs e i loro range di variazione, la fase successiva dello studio ha comportato la definizione delle Membership Functions (MFs).
La mancanza di informazioni precise circa la forma delle funzioni di appartenenza non ha reso possibile l'uso di modelli complessi, del resto sconsigliati in letteratura se non derivanti da studi mirati. L'attenzione è stata perciò rivolta a MFs lineari, triangolari o trapezoidali, le cui ripartizioni sono state fatte tenendo presente l'importanza della variabile, a cui si riferiscono, per la definizione dell'output. Ad esempio considerando la velocità del vento maggiormente vincolante rispetto alla classe di stabilità nei confronti della concentrazione di monossido, il numero di partizioni della prima sono state fatte, in prima battuta, maggiori della seconda.

2.3.2 Configurazione dei parametri del codice fuzzyTECH5.54e

Tra i metodi di allenamento previsti dal codice fuzzyTECH5.54e è stato scelto il metodo random poiché il target riguardava la quantificazione dell'errore durante la fase di validazione (vd. Cap.I).
Mentre volendo utilizzare un approccio simile al "competitive learning" di una rete puramente neurale, in quanto comporta l'uso di algoritmi di calcolo più semplici, quindi minor tempo impiegato, il numero di neuroni vincenti ("winner neurons") è stato settato al valore unitario.

In questo modo solo una regola per volta è stata modificata. I valori di Step With sono stati invece fatti variare in un range che va dal 50 al 1% (vd. Cap. I).

Inizialmente il numero massimo di iterazioni sono state fissate a 1000 e l'errore medio al 20%. Successivamente poiché interessava più il valore massimo dell'errore anche di un singolo campione rispetto al valore medio dell'intero set di dati, l'attenzione è stata rivolta all'individuazione dei valori ottimali per la Dynamic Stop Condition (vd.Cap.I)

2.3.3 Messa a punto del modello

La messa a punto del modello neuro fuzzy è stata realizzata per step successivi; l'analisi infatti è stata impostata in modo da esaminare l'influenza delle caratteristiche dei vari fattori presi singolarmente (partizioni delle MFs delle variabili, range di variazione delle stesse, metodo di defuzzificazione). Il trend degli indici di errore forniti dal codice è stato il mezzo di valutazione per l'individuazione della configurazione ottimale.

2.3.3.1 Verifica degli inputs

Tenendo presente quanto detto al §2.2.4, la prima indagine ha riguardato la verifica degli inputs da dare al codice fuzzyTECH 5.54e. Sono state svolte al riguardo numerose prove, i cui risultati hanno evidenziato come l'uso di troppi inputs, non solo comporta un maggior onere computazionale, ma addirittura possa far non convergere il codice ai limiti accettabili preimpostati (vd Tab.2.3). L'ipotesi iniziale circa la scelta degli inputs, ossia la necessità di utilizzarne una quantità inferiore rispetto a quella dell'ISC3, si è dimostrata corretta, per cui l'attenzione è stata rivolta al metodo di defuzzificazione e alla ripartizione delle FMs delle variabili.

Tab.2.3: Influenza degli inputs sugli indici di qualità del fuzzyTECH5.54e

***Inputs*(*)**	***Output*(**)**	**Errore max (%)**	**Errore medio (%)**
5	CO_{ISC3}		
sorgente, vel, dir, cs, h_m		30,38	15,34
sorgente , vel, dir, cs,T		33,56	13,9
sorgente , vel, dir, T, h_m		45,87	23,54
4			
sorgente , vel, dir, h_m		23,80	6,97
sorgente , vel, dir, cs		**19,13**	**5,42**
sorgente , vel, dir, T		26	13,2
3			
sorgente, vel, dir		24,1	15,18

(*)CO_{ISC3}= Concentrazione di monossido di carbonio calcolato mediante ISC3; vel = velocità del vento; dir= direzione del vento; cs= classe di stabilità; h_m=altezza di miscelamento;

(**) $CO_{fuzzyTECH}$=concentrazione di monossido di carbonio calcolato mediante il fuzzyTECH5.54e

2.3.3.2 Metodo di defuzzifificazione

Il fuzzyTECH5.54e permette di convertire le quantità fuzzy in valori crisp in due modi: attraverso l'applicazione del "Center of Maximum Method" (CoM) detto anche Best Compromise (BC), o attraverso l'applicazione del "Mean of Maximum Method"(MoM). Nel contesto specifico il metodo CoM si è rilevato di gran lunga il migliore ottenendo, a parità delle altre condizioni (numero di inputs, forma delle MFs, numero di ripartizioni, numero di iterazioni etc), errori massimi ed errori medi pari rispettivamente a meno della metà e a meno di un sesto di quelli ottenuti con il metodo MoM (vd Tab.2.4)

Tab.2.4: Influenza del metodo di defuzzificazione sugli indici di qualità del fuzzyTECH 5.54e

Input	***output***	***Metodo di defuzzificazione***	**Errore massimo (%)**	**Errore medio (%)**
sorgente, vel, dir,cs	CO_{ISC3}	BC	19,13	5,42
		MoM	45,54	32,24

2.3.3.3 Ripartizione delle Membership Functions

Più impegnativo è stato lo studio relativo alla ricerca delle migliori ripartizione delle MFs.
Come già detto al §2.3.1, l'idea di base è stata quella di riservare una maggiore suddivisione alle funzioni di appartenenza delle variabili che giocavano un ruolo più significativo per la determinazione dell'output. Poiché il codice implementa automaticamente un numero di regole inferiore a 1200, e non avendo una base di conoscenza onde risalire alle regole maggiormente significative, le partizioni sono state fatte rimanere limitate poiché è noto che tra le stesse e le regole esiste una stretta corrispondenza, ossia:

$$\mathrm{N}^{\circ}_{regole} = \prod_i n_i \qquad (1)$$

dove N° è il numero di regole massimo per la definizione del contesto e ni sono le ripartizioni dell'intervallo di variazione di ogni variabile.

In genere però queste non sono tutte necessarie; un indice di ridondanza delle regole può essere il parametro C(X) dato da:

$$C(X) = \sum_{k=1}^{N_r} \prod_{i=1}^{N_k} \mu_{A_{i,k}}(x_i) \qquad (2)$$

dove Nr è il numero totale di regole ed Nk il numero di ingressi coinvolti nella regola K-esima.

Se C(X)=0 la base di regole si dice incompleta, se 0<C(X)<1 sub-completa, se C(X)=1 strettamente completa, se C(X)>1 overcompleta o ridondante /9/

Nel caso studiato C(X) risulta >1 quando le ripartizioni risultano uguali a 4 per 4 variabili, mentre 0<C(X)<1 per un numero di ripartizioni inferiore.
La "completezza" della base delle regole è assicurata quindi solo nel caso in cui tutti gli inputs, ad eccezione della classe di stabilità, e l'output abbiano il numero di ripartizioni delle MFs pari a 4 (vd Tab. 2.5). Questo assicura l'assenza di non linearità tipo "isteresi" nell'algoritmo di inferenza /9/ e da notare è come gli errori massimo e medio forniti dal codice rimangano in tal caso ben al di sotto della soglia di accettabilità (20%).

Tab.2.5: Influenza delle ripartizioni delle MFs sugli indici di qualità del fuzzyTECH 5.54e

Input	*output*	*Ripartizioni delle MFs*		Errore massimo (%)	Errore medio (%)	C(X)	
		3	*4*			>1	<1
CO_{ISC3}, vel, dir,cs	$CO_{fuzzyTECH}$	CO_{ISC3}, vel, dir,cs, sorgente	-	19,13	5,42	no	si
		CO_{ISC3}, vel, dir, cs	sorgente	17,18	5,7	no	si
		CO_{ISC3}, vel, cs, sorgente	dir	15,41	4,9	no	si
		dir, cs, CO_{ISC3} e sorgente	vel	18,32	6,3	no	si
		dir, vel, cs, sorgente	CO_{ISC3}	19,15	8,4	no	si
		vel, cs, sorgente	CO_{ISC3}, dir	18,76	6,29	no	si
		dir,cs sorgente	CO_{ISC3}, vel	13,14	5,2	no	si
		vel, dir,cs	CO_{ISC3} sorgente	15,09	4,37	no	si
		sorgente, cs, CO_{ISC3}	dir, vel	10,34	3,6	no	si
		vel, cs CO_{ISC3}	dir sorgente	15,72	3,67	no	si
		dir, cs CO_{ISC3}	vel sorgente	14,33	11,8	no	si
		cs, sorgente	dir, vel, CO_{ISC3}	6,08	3,57	no	si
		cs, CO_{ISC3}	dir, vel, sorgente	4,35	4,01	no	si
		cs, dir	vel, CO_{ISC3} sorgente	7,19	6,03	no	si
		cs, vel	CO_{ISC3} sorgente dir	9,2	5,84	no	si
		cs	**CO_{ISC3} sorgente dir, vel**	**3,4**	**1,3**	**si**	no

2.3.3.4 Range di variazione dell'output

Un passo ulteriore nella definizione del modello ottimale ha riguardato l'analisi dell'influenza dei range di variazione degli inputs e dell'output. Mentre per i primi l'allargamento degli estremi delle MFs su cui far operare il codice porta a miglioramenti non significativi, se non addirittura a condizioni peggiori, per l'output un sensibile spostamento dei limiti, tale da aumentare il set di dati di inferenza, porta ad un miglioramento apprezzabile riscontrabile in entrambi gli indici di qualità e pari a più del 40% (vd.Tab.2.6)

Tab.2.6: Influenza del range dell'output sugli indici di qualità del fuzzyTECH 5.54e

Input	***output***	***Range di variazione***	**Errore massimo (%)**	**Errore medio (%)**
sorgente, vel, dir,cs	CO_{ISC3}	$0 \leq CO_{ISC3} \leq 100$	3,4	1,3
		$0 \leq CO_{ISC3} \leq 130$	1,98	0,76

2.4 Verifica della potenziale capacità dello strumento di ottenere risultati paragonabili ad una metodologia già validata

Utilizzando il modello ritenuto migliore, avendo indici di qualità della fase di training notevolmente superiori (errori più piccoli) rispetto agli altri, si è proceduto con la fase di validation.
Estratti inizialmente in modo del tutto casuale dal set completo un terzo dei dati generati da ISC3 e non usati durante la fase di allenamento, sono stati impiegati i valori di inputs di questi per calcolare la concentrazione di monossido di carbonio mediante i pesi delle regole selezionate nella precedente fase di training (vd App.3).

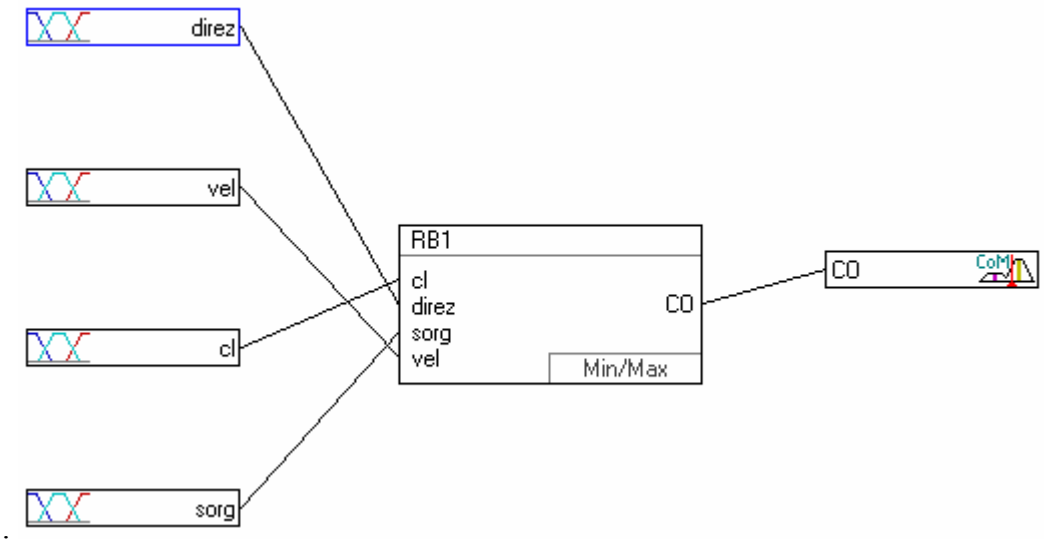

Fig.2.2: Modello del codice fuzzyTECH5.54e

Quanto ottenuto è riportato in App.3 (b)

Con tali valori di concentrazione non è risultato possibile applicare il criterio di accettabilità stabilito al §2.2.1 per la presenza di un numero cospicuo di zeri.
Si sono perciò analizzati i trend dei due output (vd. Figg.2.3-2.4) ed è stato calcolato il *Pearson product-moment correlation coefficient (r)*

Come risulta evidente dalla Fig.2.3, i dati calcolati dal fuzzyTECH5.54e hanno un andamento pressoché sovrapponibile a quelli di ISC3.

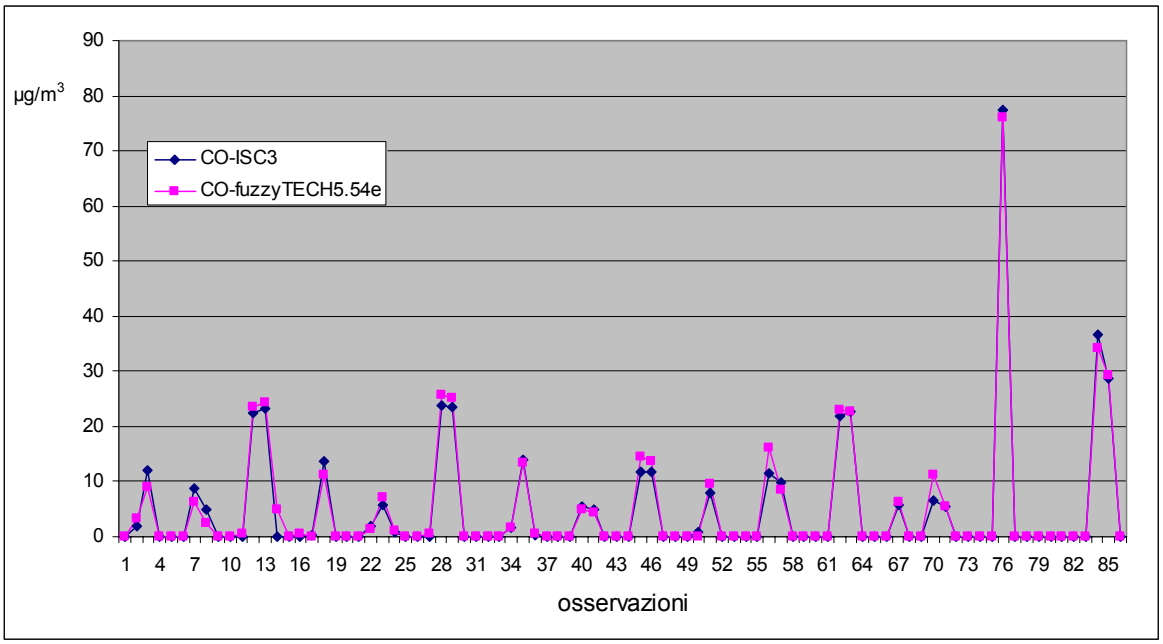

Fig.2.3: Concentrazione di monossido di carbonio calcolato da ISC3 e dal fuzzyTECH5.54e

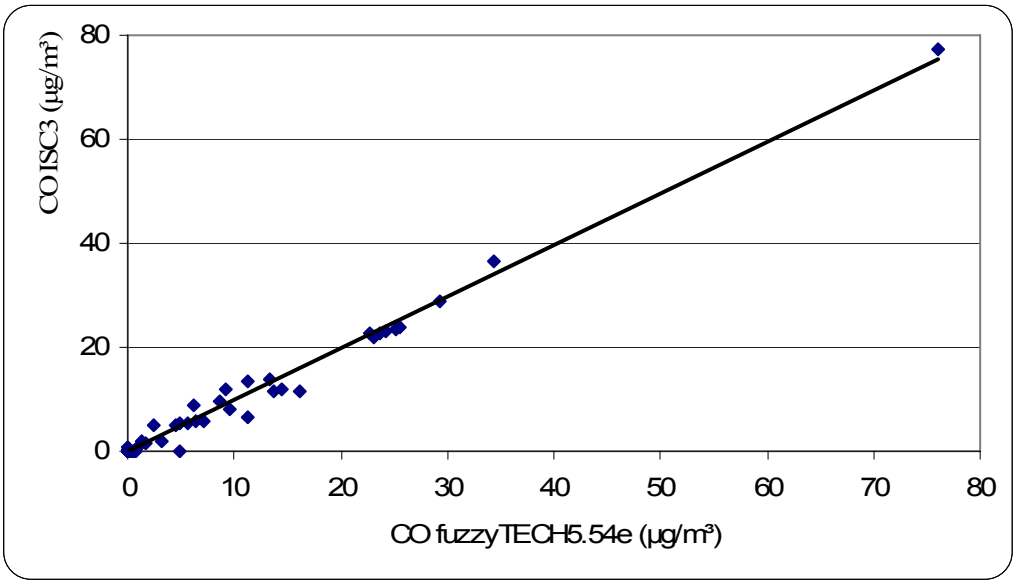

Fig.2.4: CO di ISC3 versus CO del fuzzyTECH5.54e

Inoltre l'indice di correlazione risulta pari a 0,994 a conferma della stretta somiglianza esistente tra le due

variabili

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right) * \left(y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2 * \sum_{i=1}^{n}\left(y_i - \bar{Y}\right)^2}} = 0,993693$$

dove x_i e y_i sono i valori assunti dalla variabile x e dalla variabile y all'istante i, mentre $\bar{X}$ e $\bar{Y}$ sono le medie delle rispettive popolazioni

2.5 Conclusioni

Quanto sopra riportato porta a concludere che una rete neuro-fuzzy riesce ad avere, nelle stesse condizioni, almeno un grado di affidabilità pari a quello di un codice classico.

Il valore dell'indice r, calcolato al §2.4, assicura una correlazione talmente stretta tra le due quantità da farle ritenere rappresentanti della stessa distribuzione di dati /10/.

III. Applicazione al caso in esame

3.1 Premessa

Avendo appurato la validità della metodologia ibrida nell'ambito della valutazione della qualità dell'aria in un contesto urbano (vd cap.II), la fase successiva dello studio è stata mirata alla definizione delle linee guida da adottare nell'analisi diagnostica e prognostica dell'inquinamento ambientale da fonte veicolare utilizzando tecniche neuro-fuzzy, nel caso in cui le caratteristiche meteorologiche della città esaminata (es. brezze marine, effetti canyon etc), associate ad un profilo urbanistico complesso che alterna ampi viali a strade strette, rendano non possibile l'uso di codici gaussiani.

3.1.1 Inquinante da indagare

La scelta di cosa indagare è ricaduta sul CO fondamentalmente per due motivi:

- la quantità emessa dagli autoveicoli è circa 10 volte maggiore di quella degli altri inquinanti (nell'emisfero nord la concentrazione è di circa 0,1-0,2ppm; nell'emisfero sud 0,04-0,06ppm) e per la sua scarsa reattività viene spesso utilizzato come tracciante dell'andamento temporale degli inquinanti a livello del suolo[1];
- gli effetti sull'uomo possono essere particolarmente gravosi

Il Monossido di carbonio si forma ogni volta che sostanze contenenti carbonio sono bruciate in difetto d'aria, tuttavia, anche quando la quantità è teoricamente sufficiente, le reazioni non procedono fino al completamento poiché quella di formazione del CO è dieci volte maggiore di quella di ossidazione del CO a CO_2

Una volta inalato, il CO si combina con l'emoglobina del sangue e formando la carbossiemoglobina provoca l'anossia cellulare, in particolare del cervello e del cuore (la sua affinità con l'emoglobina è 220 volte superiore a quella dell'ossigeno). La sua emivita varia da 4 a 6 ore dipendendo dal grado d'attività fisica del soggetto, dalle richieste d'ossigeno da parte dei tessuti e dalla concentrazione d'emoglobina. A concentrazioni dell'ordine dell'0,01% il soggetto esposto non manifesta generalmente sintomi d'intossicazione. A concentrazioni dell'ordine dell' 0,05%, in corso d'attività fisica moderata, possono manifestarsi segni di cefalea leggera. Un grado d'esposizione prolungata o attività fisiche più impegnative causano cefalea, irritabilità, stato confusionale, disturbi visivi, stordimento, nausea e vomito. Dopo l'esposizione per un'ora a concentrazioni di 0,1% nell'aria inspirata il sangue contiene il 50 – 80% di carbossiemoglobina che produce convulsioni, insufficienza respiratoria, coma e morte. L'inalazione di concentrazioni maggiori di CO satura l'emoglobina così rapidamente che la perdita della coscienza può avvenire improvvisamente e senza segni premonitori.

Il segno più tipico di una grave intossicazione da CO è una colorazione caratteristica della cute e delle mucose, dovuta al colore rosso brillante della carbossiemoglobina.

3.1.2 Contesto di analisi

Il contesto preso a riferimento è stato quello livornese poiché rappresenta un'area urbana mediamente popolata (circa 156.000 abitanti al censimento del 2003), dove il contributo del traffico veicolare costituisce una sorgente non trascurabile rispetto alle altre fonti d'inquinamento e la sua posizione geografica fa sì da conferirgli le caratteristiche dette nella premessa.

Sul territorio sono attive due reti di monitoraggio indipendenti. La rete provinciale, gestita dall'ARPAT territorialmente competente, è costituita da 5 centraline localizzate sul territorio in modo tale da poter ottenere un quadro significativo, sia del livello d'inquinamento, sia, quando possibile, di informazioni utili per individuare la causa dei valori registrati. Tutte le centraline sono dotate di strumentazione per la misurazione delle concentrazioni degli inquinanti previsti dalla normativa. In corrispondenza delle centraline di Piazza Mazzini e Viale Carducci è stata posta una telecamera, dotata di fotocellula, in grado di registrare il volume di traffico orario, suddividendo i veicoli in classi in funzione della loro lunghezza

[1] il CO rimane in atmosfera per circa 3 o 4 mesi e si rimuove per ossidazione a CO_2 o per reazioni fotochimiche coinvolgenti il metano e i radicali OH

Fig. 3.1: Centralina e fotocellula lungo Viale Carducci

L'altra rete di monitoraggio, rete ARIAL, gestita dall'Associazione fra gli Industriali della Provincia di Livorno, è costituita da cinque centraline poste nella zona settentrionale del centro urbano più una sesta stazione (in via Marx), situata in zona industriale, nel comune di Collesalvetti. Analogamente a quanto avviene per le centraline della rete provinciale, l'analisi dei campioni d'aria prelevati riguarda gli inquinanti previsti dalla normativa vigente.
La relativa semplicità di reperimento dei dati necessari per lo svolgimento dello studio ha fatto preferire V.le Carducci alle altre realtà locali

Fig.3.2: Viale Carducci

3.1.3 Acquisizione dei dati

I dati acquisiti hanno riguardato i mesi di Giugno e Luglio dell'anno 2002, in particolare:

- i dati meteorologici registrati dalla stazione La Rotonda Ardenza gestita dall'ARPAT ;
- il traffico autoveicolare medio orario registrato dalla fotocamera posta sulla centralina di V.le Carducci (fonte ARPAT);
- la concentrazione di monossido registrata dalla centralina posta su V.le Carducci (fonte ARPAT).

3.2 Analisi preliminare

La definizione delle linee guida è stata determinata partendo dalla messa a punto di un modello neuro-fuzzy capace di effettuare l'analisi diagnostica e prognostica della concentrazione di CO lungo l'arteria presa a riferimento.
A tal fine sono state individuate, tenendo presente anche quanto acquisito durante la fase di assessment e mediante uno studio morfologico (area interessata a brezze; area collinare; area delimitata da catene montuose etc.), climatico (temperatura, pressione, umidità, radiazione solare, altezza di miscelamento, classi di stabilità etc.), dell'uso del suolo e del tipo di trasporti, le variabili maggiormente significative per la definizione del contesto (Tab.3.1).

Tab.3.1:Possibili variabili necessarie per la caratterizzazione della realtà locale

Input	Output
Direzione del vento Velocità del vento Classe di stabilità Flusso autoveicolare Fondo di Monossido	Concentrazione di monossido registrato dalla centralina

3.2.1 Il fondo

Dall'analisi dell'uso del suolo è emerso che le sorgenti di monossido di carbonio, oltre al traffico veicolare, potevano essere le attività antropiche ivi esistenti (nell'area è presente un insediamento industriale e un porto significativi a livello nazionale) e le fonti di riscaldamento (stufe a gas, a legna, a olio combustibile). Tuttavia svolgendo lo studio durante un periodo estivo il contributo delle fonti di riscaldamento è stato ritenuto trascurabile e il fondo è stato interamente attribuito alle attività industriali, al porto e alle arterie stradali presenti sul sito a prescindere da V.le Carducci. Lo stesso è stato perciò calcolato utilizzando ISC3st, al quale sono stati presentati i dati meteo del periodo esaminato e nel quale sono state "selezionate" le sorgenti specifiche del caso.
Una volta ottenuti i valori delle concentrazioni di monossido di carbonio per i mesi di Giugno e Luglio relativi alle attività antropiche dell'area, si è proceduto ad una loro analisi onde poter evidenziare eventuali incongruenze.
Analizzando il trend rispetto alla concentrazione misurata su V.le Carducci, non sono emerse circostanze "non accettabili".
Il fondo è sempre rimasto inferiore alla concentrazione rilevata dalla centralina e il suo valore è sempre diminuito all'aumentare della velocità del vento (vd All.4)
Inoltre dall'analisi del layout del sito (vd Fig.3.3), il contributo dell'area industriale e del porto interessano maggiormente V.le Carducci se la direzione del vento rimane nel range compreso tra 75° e 135°.

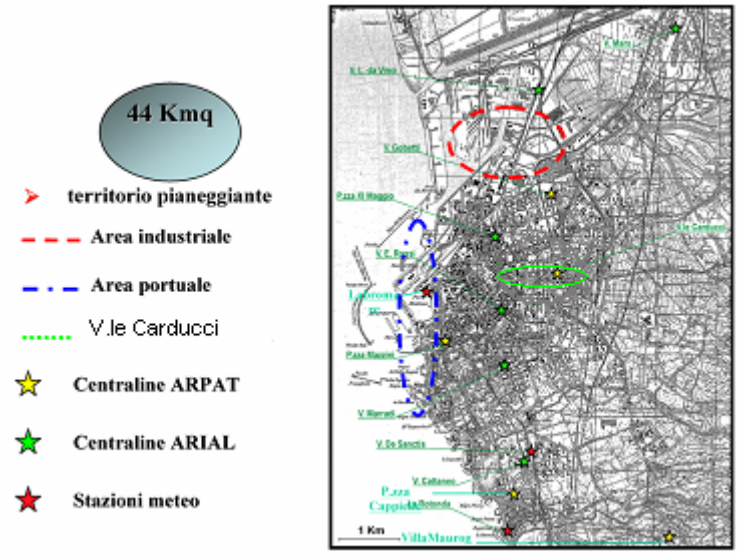

Fig.3.3: Posizione di V.le Carducci rispetto all'area industriale e al porto

Per vedere se tale andamento si fosse verificato anche per i dati calcolati, dando un elemento in più per accertare la loro validità, è stato deciso di impiegare un indice statistico che rappresentasse meglio i trend della direzione del vento e del fondo eliminando, anche parzialmente, le fluttuazioni stocastiche dei dati. La preferenza è ricaduta sulla media mobile con un numero di periodi pari a sette. Tale scelta è scaturita dalla necessità di ottenere un buon compromesso tra precisione dell'andamento delle variabili e perdita di informazioni sulle variazioni a breve intervallo di tempo.
Come mostrano le Figg. 3.4 e 3.5 il fondo tende ad aumentare in modo marcato (da valori dell'ordine di 1 $\mu g/m^3$ può arrivare anche a 35 $\mu g/m^3$e oltre cioè un incremento di più del 3000%) quando la direzione del vento rimane compresa tra 75° e 135°, confermando pienamente quanto trovato dall'analisi del sito.
L'analisi degli andamenti puntuali (vd. App.4) ripetono quanto già detto per l'87% dei casi; ciò ha portato a concludere che le fluttuazioni presenti non influenzano pesantemente la validità del set di dati per ciò che concerne l'andamento del fondo in relazione alla direzione del vento.

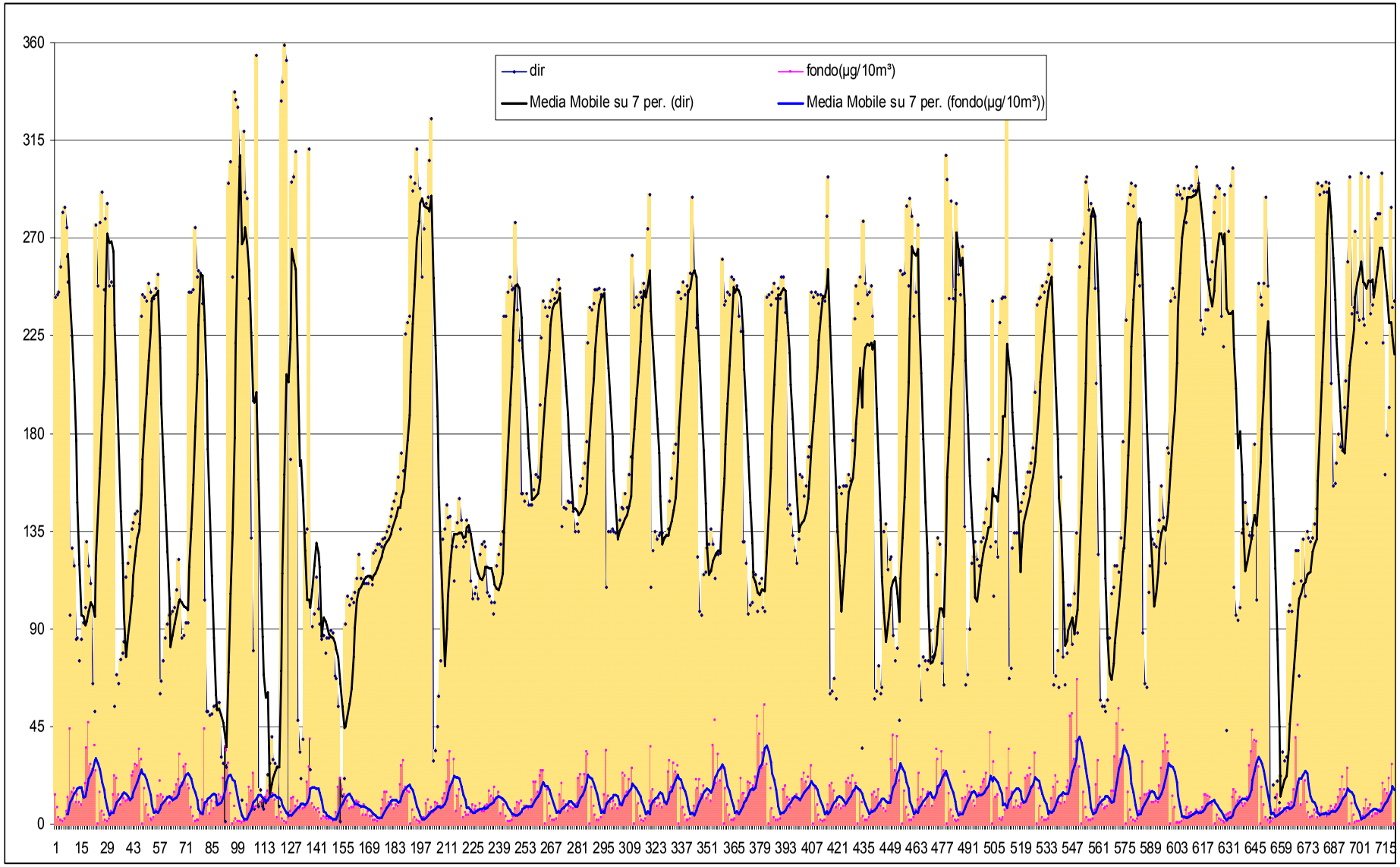

Fig.3.4:Andamento della direzione del vento e del fondo per il mese di Giugno

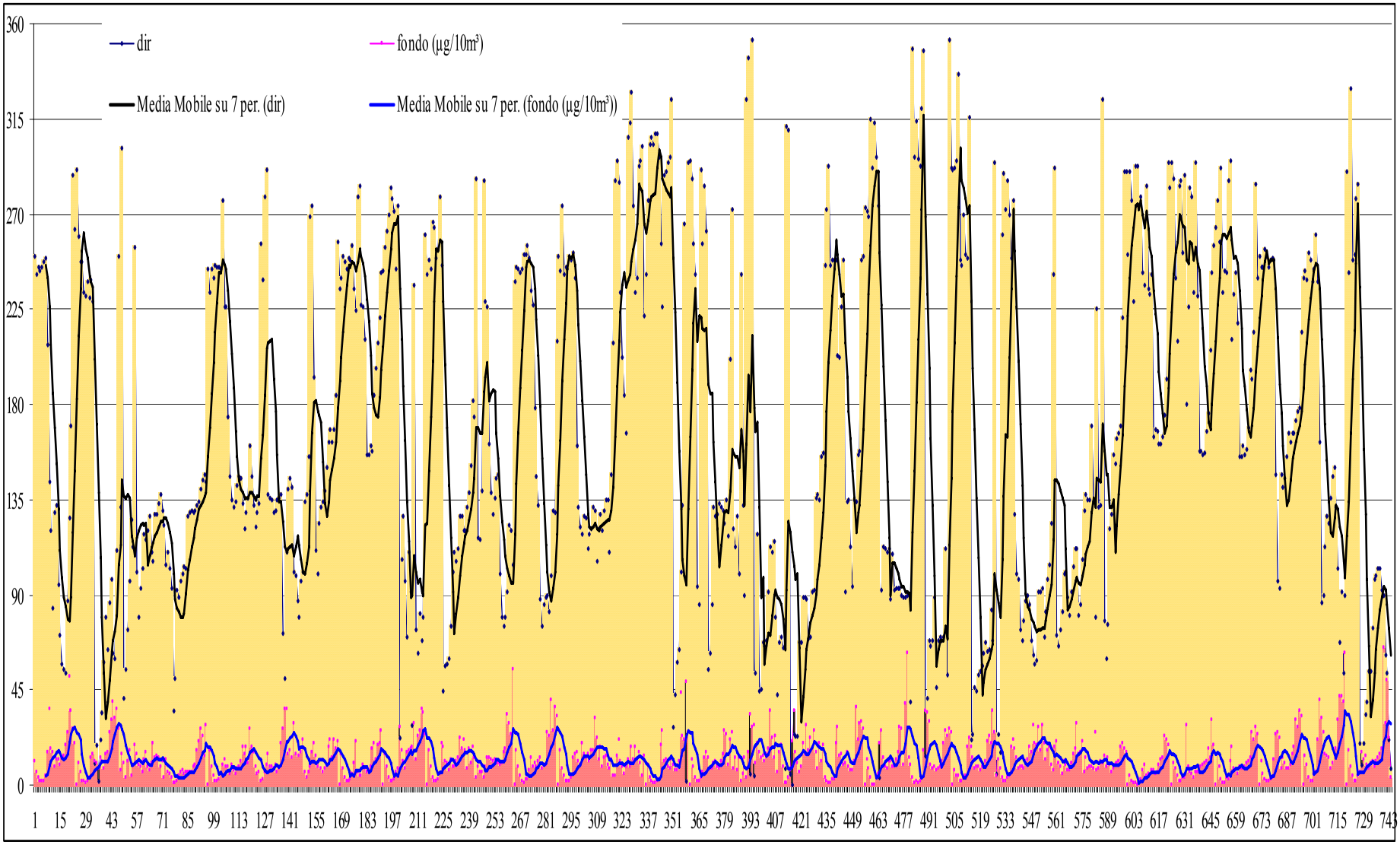

Fig.3.5: Andamento della direzione del vento e del fondo per il mese di Luglio

3.2.2 Variabili utili alla definizione del contesto

Successivamente è stata svolta una ricerca per rilevare eventuali correlazioni tra le variabili d'ingresso, in modo da eliminare probabili ridondanze causa di un maggior onere computazionale. Utilizzando il *Pearson product-moment correlation coefficient (r)* (vd. §2.4), è stato ottenuto quanto riportato in Tab.3.2

Tab.3.2**:**Correlazione tra le variabili d'ingresso

r	**dir**	**cs**	**Fondo**	**vel**	**veic**
dir	**-**	**-**	**-**	**-**	**-**
cs	0,07	-	-	-	**-**
Fondo	-0,46	0,01	-	-	**-**
vel	0,00	0,18	0,38	-	**-**
veic	0,00	0,00	0,001	0,00	**-**

Poiché $|r| \leq 0{,}8$ le variabili possono essere ritenute non correlabili(*), tutte verosimilmente utili alla definizione del contesto.

3.3 Analisi dei dati

Una volta appurata la necessità di utilizzare le variabili riportate al paragrafo precedente per creare un modello neuro-fuzzy diagnostico e prognostico dell'inquinamento da CO, la difficoltà incontrata è stata quella di muoversi in uno spazio multidimensionale; ogni osservazione infatti è caratterizzata da 6 valori (5 input, 1 output) che possono essere trattati alla stregua di una matrice; poiché per ottenere risultati attendibili le fasi di training e di validation devono avvenire con dati che appartengono ad una stessa regione di spazio, il primo target è stato quello di individuare la disposizione spaziale di tali vettori.

3.3.1 Analisi bidimensionale delle variabili di ogni mese e loro raffronto

Si è pensato in prima battuta di mappare tutte le variabili a coppie per confrontarne l'andamento nei due mesi disponibili e, quando possibile perché fisicamente significativo, ricercarne una possibile correlazione. In questo modo è stato possibile verificare anche la validità dei dati esaminando la dispersione degli stessi in un piano bidimensionale.
Non sono emersi al proposito casi di particolare evidenza: generalmente i dati sono rimasti distribuiti in spazi relativamente contenuti e solo per alcuni è stato possibile l'eliminazione diretta perché mostravano essere "troppo distanti" dagli altri .
Inoltre, ipotizzando l'assenza di grandi errori di misura e ritenendo i mesi considerati paragonabili da un punto di vista di traffico urbano e meteorologico, gli andamenti sarebbero dovuti essere similari. Tale ipotesi è risultata pienamente verificata per quanto riguarda la coppia di variabili "traffico veicolare orario" e "concentrazione di monossido di carbonio" (COt) misurata dalla centralina di V.le Carducci (vd.App.4)
Non altrettanto però è stato possibile dire per le altre variabili.
Per quanto riguarda la "direzione del vento" e la "concentrazione di monossido di carbonio" non si è riusciti a formulare alcuna ipotesi per la grande variazione della direzione del vento (vd. App.4). Ciò che si è potuto notare è che perde di significato anche la possibilità di una relazione tra aumento della concentrazione di CO e andamento della direzione, come accadeva invece per il fondo (vd. § 3.2.1).

Effettivamente risultando il profilo degli edifici disomogeneo e con frequente presenza di rientranze dovuto al non allineamento dei palazzi lungo la strada (vd.Fig.3.6a), potrebbero verificarsi sia un incanalamento del vento all'interno del tratto stradale che vortici secondari tali da far venir meno la dipendenza della

concentrazione registrata dalla centralina e la direzione del vento fornita dalla stazione meteo posta in una zona poco urbanizzata (vd Fig.3.6b).

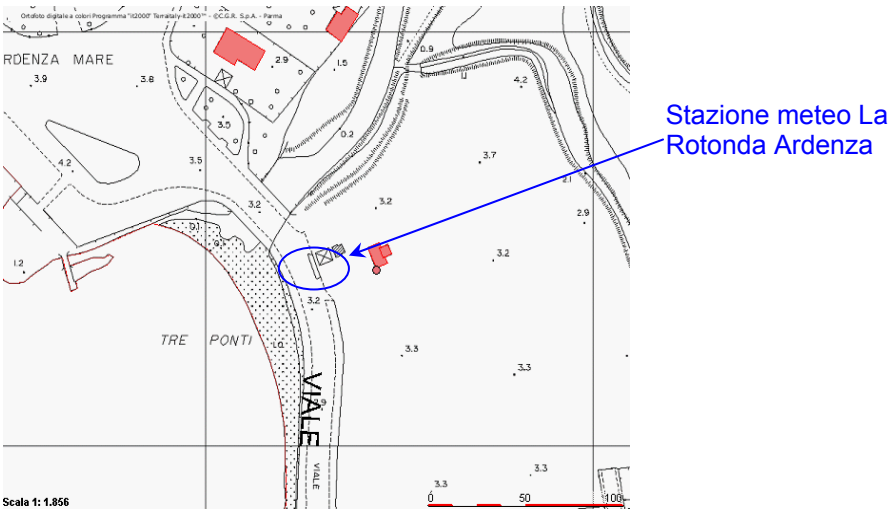

Fig.3.6(a): Urbanizzazione del sito dove è presente la stazione meteo'La Rotonda Ardenza'

Fig.3.6(b): Urbanizzazione del sito dove è presente la centralina di V.le Carducci

Per quanto riguarda la "velocità del vento" e la "concentrazione di monossido di carbonio" è stato possibile confermare la dipendenza inversa tra le variabili per entrambi i mesi. La velocità rimane generalmente contenuta entro valori prossimi a 8m/s (meno dell'1,4% per giugno e meno del 6‰ per luglio superano tale soglia) e solo il 3% del totale supera i 6m/s (5% per giugno e l'1% per luglio). Non risultano tuttavia equiparabili in prima istanza gli andamenti corrispondenti; spesso infatti la crescenza della velocità per un set di dati di un mese non corrisponde agli stessi del mese successivo dove o l'ordine di grandezza è nettamente diverso o addirittura il trend può essere inverso (vd. App.4).

L'analisi degli andamenti della "classe di stabilità" e della "concentrazione di monossido di carbonio totale", nonché della "classe di stabilità" e della "concentrazione del fondo" ha evidenziato una dipendenza tra le variabili coerente con la turbolenza associata alla classe di stabilità.

Si può notare infatti che, in entrambi i casi, le concentrazioni risultano maggiori per classi di stabilità A e B e decrescono al procedere di questa verso la F.
I due mesi risultano del tutto similari: le classi E ed F si verificano sempre durante le ore notturne e durante le prime ore del mattino; le concentrazioni di monossido aumentano in fasce orarie ben definite, dalle 7 alle 10, dalle 12 alle 14 e dalle 18 alle 22. (vd. App.4) mentre il fondo cresce nelle ore centrali della giornata (vd. App.4)
L'analisi della coppia di variabili "classe di stabilità" e "velocità del vento" ha evidenziato invece alcune contraddittorietà: in entrambi i mesi più del 10% dei dati della classe di stabilità non risulta avere una velocità compatibile alla sua classificazione; l'incongruenza risulta maggiore per velocità superiori o uguali a 4m/s, anche se numericamente poco significative rispetto al set completo di dati (< 4%).
Tuttavia l'andamento della classe di stabilità mostra essere abbastanza similare per Giugno e Luglio, altrettanto non è possibile affermarlo per la velocità. Conseguentemente non è stato possibile fare un confronto a coppie (vd. App.4).
Infine dall'esame delle rimanenti coppie di variabili ("direzione del vento-velocità del vento"; "direzione del vento - classe di stabilità"; "direzione del vento-flusso veicolare"; "velocità del vento-flusso veicolare; "flusso veicolare -fondo di monossido di carbonio"; "classe di stabilità-flusso veicolare") non si evincono particolari similitudini tali da costituire una base di conoscenza per la disposizione spaziale delle osservazioni (vd. App.4).

3.3.2 Rappresentazione grafica degli andamenti di coppie di variabili

Non avendo ottenuto particolari indicazioni dallo studio svolto al § 3.3.1, si è pensato di procedere attraverso la sovrapposizione degli andamenti di coppie di variabili corrispondenti per i mesi esaminati. Se un mese avesse sempre contenuto per ogni coppia di variabili quelle corrispondenti dell'altro mese, avremmo potuto stabilire che il primo mese conteneva spazialmente il secondo e ciò avrebbe portato a concludere che i primi dati potevano costituire un "inviluppo" dei secondi. Ergo l'utilizzo del "set comprendente" come dati di allenamento e l'utilizzo del "set compreso" come dati di validazione.
Non è stato possibile però definire la prevalenza di un mese rispetto all'altro: per la coppia "fondo-velocità del vento", infatti il mese di Giugno è sembrato predominare su Luglio, negli altri casi l'andamento è stato così irregolare da non poter formulare alcuna ipotesi (vd Fig.3.7)

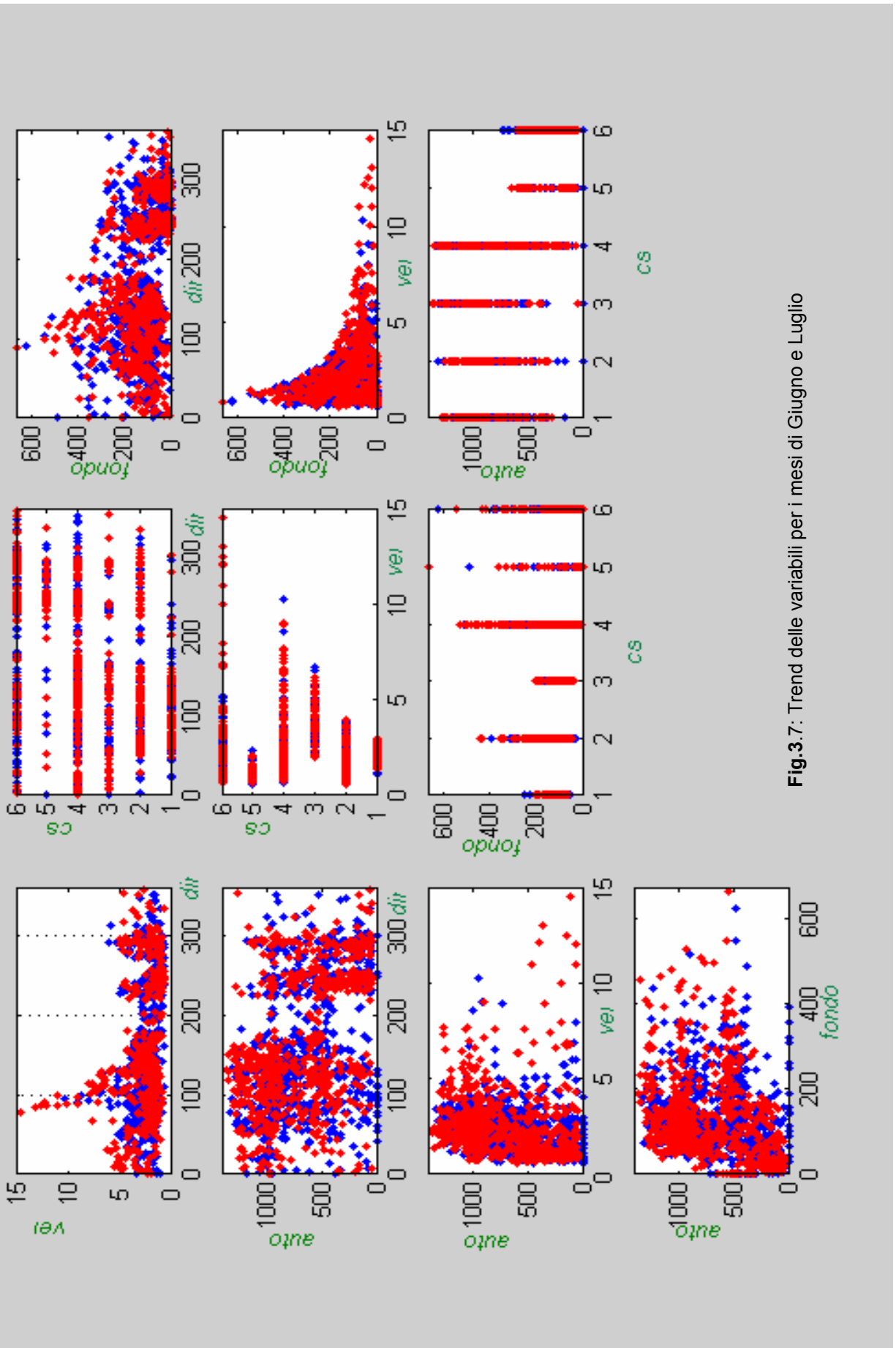

Fig.3.7: Trend delle variabili per i mesi di Giugno e Luglio

3.3.3 Analisi cluster

L'analisi degli andamenti puntuali e il confronto a coppie dei dati svolti al §3.3 hanno confermato in linea di massima la validità dei dati, ma niente è stato possibile dedurre circa la posizione relativa degli stessi. E' emersa a questo punto l'esigenza di effettuare un'indagine mirata in modo da individuare, quando possibile, gruppi di dati "simili". Tale valutazione è stata motivata dalla necessità di:

- eliminare eventuali ridondanze e/o conflitti non emersi in precedenza che avrebbero comportato solo un maggior onere computazionale;
- selezionare solo i dati strettamente necessari per le informazioni del caso.

A tal fine è stata affrontata un'analisi cluster utilizzando inizialmente un metodo gerarchico classico /8/, denominato ISODATA.
Impiegando valori diversi dell'indice di accuratezza[2], sono stati selezionati 10 gruppi di riferimento di dati variabili tra 1200 e 120. Questi ultimi sono stati sfruttati per la fase di allenamento del codice fuzzyTECH5.54e, ma gli indici di qualità che sono emersi non sono stati particolarmente soddisfacenti (errore massimo compreso tra il 20 e il 30% ed errore medio tra il 7 e il 10%; vd Fig.3.8) se confrontati a quelli ottenuti durante la fase di assessment (vd § 2.3.3).

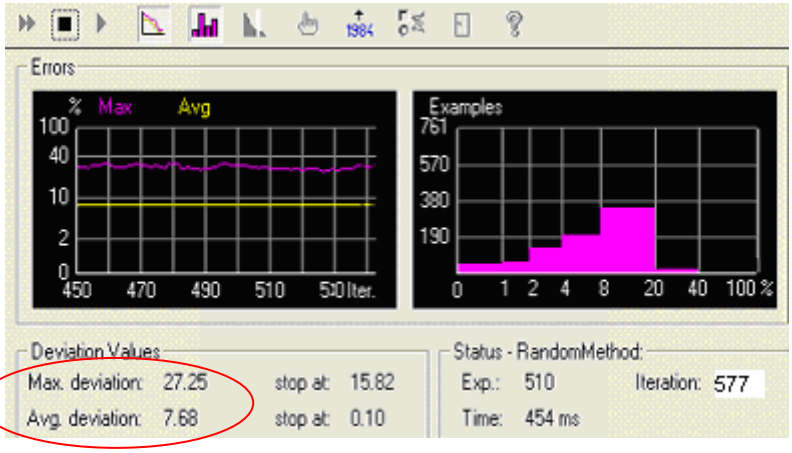

Fig.3.8**:** Indici di qualità del fuzzyTECH5.54e dopo il training con i dati clusterizzati da Isodata

Si è ipotizzato che l'approccio classico, mediando le distanze su tutte le variabili, facesse perdere troppe informazioni e non fosse perciò quello ideale per il caso specifico.
L'attenzione è stata allora rivolta ad un metodo di clusterizzazione fuzzy (FuzzyCluster /10/), dove distanze ed accuratezza vengono trattate separatamente per ogni variabile. L'indice di accuratezza viene definito adesso tramite due parametri, ε e δ, che rappresentano rispettivamente la massima differenza espressa in percentuale tra due punti che possono essere ritenuti simili e la più piccola differenza in % tra due punti che non possono essere ritenuti più simili (Fig.3.9).
Se due punti hanno una differenza in % compresa nell'intervallo [-ε , -δ] o nell'intervallo [ε , δ] essi possono essere ritenuti simili con un grado di appartenenza espresso dalla membership function.
In questo modo la procedura di clusterizzazione viene eseguita in due fasi. Dopo il calcolo delle distanze i dati campione con il più alto "grado di verità", espresso in questo caso da un vettore a sei dimensioni, sono considerati "chiusi" e clusterizzati in un nuovo dato.

[2] Rappresenta la massima distanza, espressa in %, che può intercorrere tra due set di dati ritenuti simili

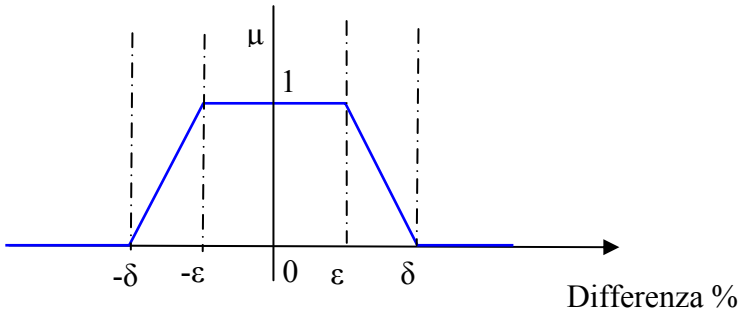

Fig.3.9: Criterio di similitudine tra due punti utilizzando una Membership Function (FuzzyCluster)

Utilizzando un simile approccio sono stati ricavati set di dati di numerosità variabile (da poche centinaia a poco più di un migliaio) dipendente dal grado di accuratezza preimpostato. E' emerso che il gruppo di dati migliori (993 dei 1400 iniziali), cioè quelli che fanno ottenere dopo l'allenamento del codice indici di qualità più vantaggiosi, hanno il parametro δ compreso tra 5 e 7 ed ε compreso tra 2 e 3. Per valori di δ e di ε maggiori il set di dati ottenuto non risulta sufficientemente significativo per riprodurre, entro un margine di errore accettabile, tutti i casi contemplati nell'insieme iniziale.
L'impiego del FuzzyCluster ha migliorato sicuramente il grado di precisione nella fase di training (errore massimo minore del 20% ed errore medio minore del 7 %; vd Fig.3.10), ma tale progresso non è stato tale da poter costruire un modello neuro-fuzzy attendibile. Gli errori infatti durante la validation del codice aumentano significativamente (errore massimo dell'ordine del 40% ed errore medio dell'ordine del 30 %) e sarebbero quindi risultati inaccettabili se pensati impiegati in una fase prognostica.

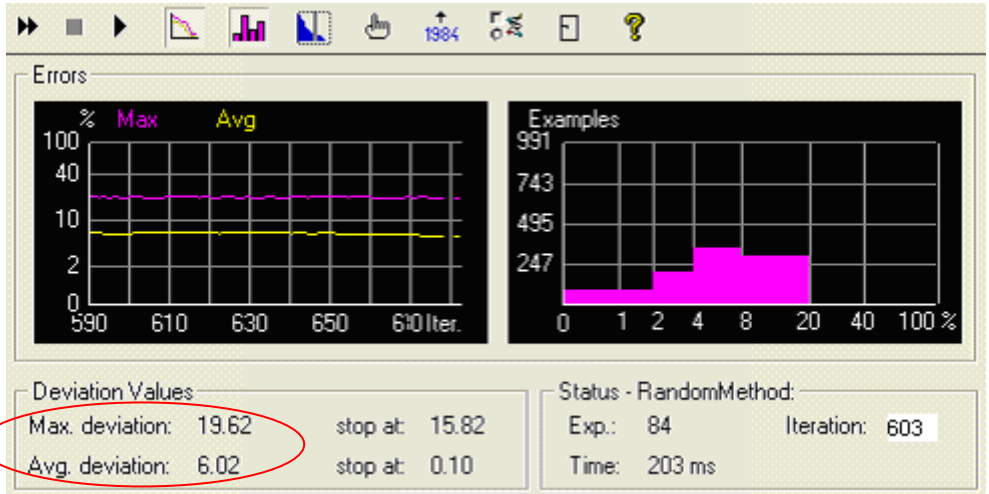

Fig.3.10**:** Indici di qualità del fuzzyTECH5.54e dopo il training con i dati clusterizzati da FuzzyCluster

3.3.4 Modelli neuro-fuzzy

Poiché l'analisi cluster non ha portato all'individuazione di un set di dati ottimali per l'allenamento del codice si è pensato di utilizzare gli stessi integralmente per verificare in quale misura le informazioni erano contenute nell'insieme iniziale e, se possibile, ricavare un ordine di grandezza dell'errore associato ad una possibile loro ripartizione.
Con tali presupposti è stato individuato un modello (vd.App.5) dove le partizioni delle funzioni di appartenenza (vd Cap.II) sono state assegnate tenendo presente sia la rilevanza della singola variabile di input per la determinazione dell'output, sia il suo range di variazione, sia il limite imposto dal codice (numero di regole < 1200_vd.§ 2.3.3.3).
Come mostra la Fig.3.11 l'errore massimo decresce fino a circa il 15%, a conferma del fatto che la riduzione del set iniziale utilizzando metodiche di clusterizzazione non apporta alcun vantaggio. Ripetendo lo studio utilizzando i dati mensili in modo separato (vd App.6) è emerso un netto peggioramento degli indici di qualità (vd. Figg.12-13).
Si è ipotizzato a questo punto che i dati a disposizione non fossero uniformemente distribuiti nello spazio, ipotesi certamente plausibile dal momento che l'area interessata è esposta a brezze marine e le caratteristiche urbanistiche favoriscono la creazione di vortici locali.

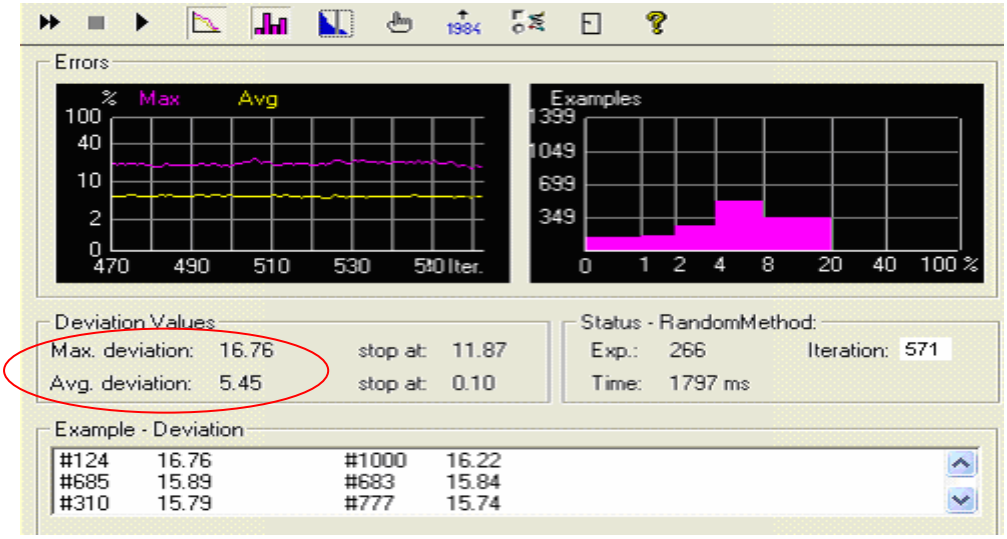

Fig.3.11**:** Indici di qualità del fuzzyTECH5.54e utilizzando il set completo di dati

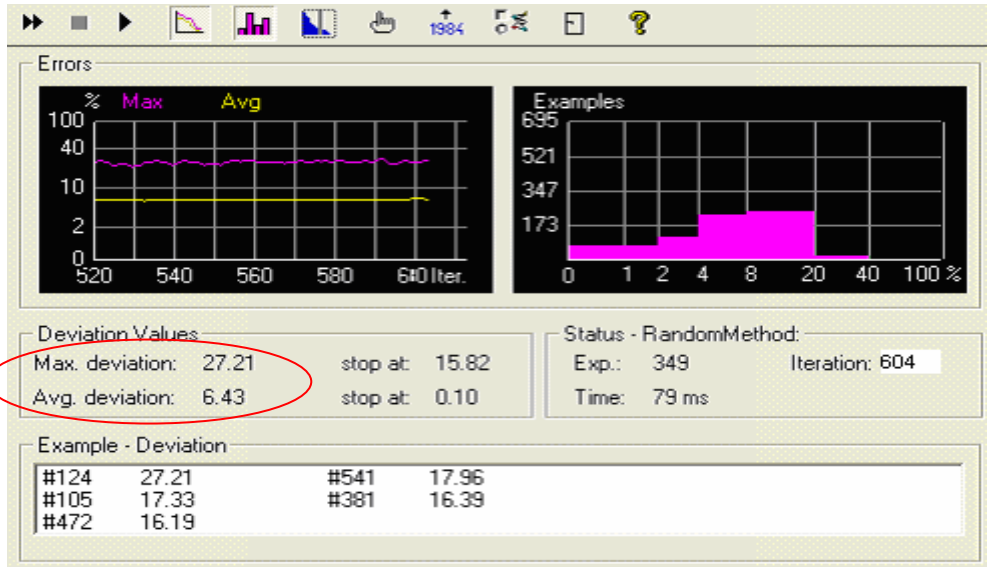

Fig.3.12**:** Indici di qualità del fuzzyTECH5.54e utilizzando i dati di giugno

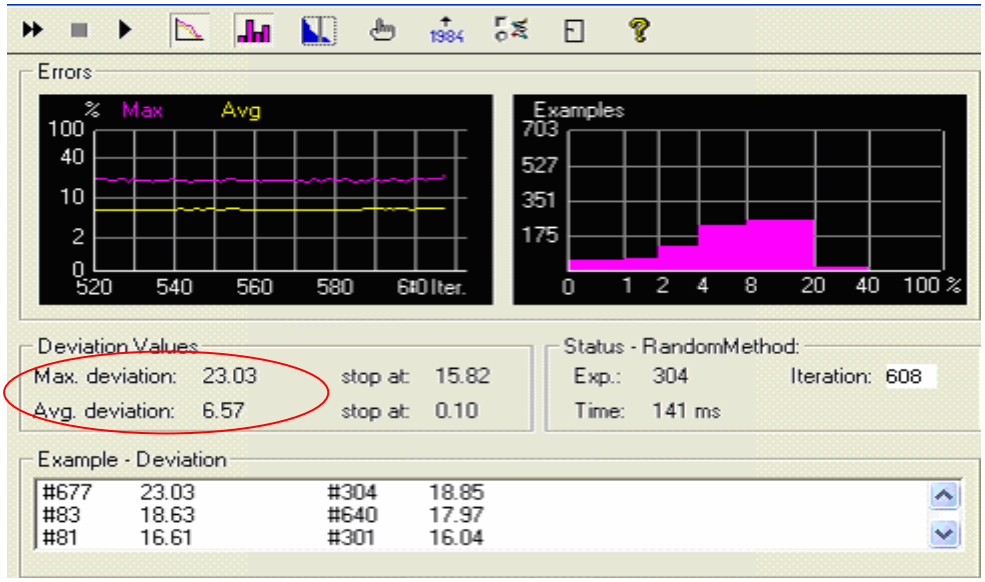

Fig.3.13**:** Indici di qualità del fuzzyTECH5.54e utilizzando i dati di luglio

3.4 Studio delle regole fuzzy

Utilizzando integralmente o mensilmente i dati non è stato possibile ottenere modelli neuro fuzzy affidabili per l'analisi prognostica dell'inquinamento (vd § 3.4); l'ipotesi è stata quella di essere in presenza di una distribuzione spaziale non uniforme degli stessi ed è emersa a questo punto la necessità di eliminare eventuali mancanze di continuità restringendo l'analisi a quegli spazi effettivamente coperti dai dati e rimuovendo possibili border line. Poiché le analisi cluster svolte precedentemente non hanno apportato informazioni significative si è pensato di iniziare dallo studio delle regole fuzzy.

3.4.1 Analisi delle regole fuzzy

Lo studio è stato svolto in tre fasi:

- una prima analisi volta alla determinazione della prevalenza dei dati partendo dalla ripartizione della direzione in 12 settori effettuata mediante la divisione della rosa dei venti in gruppi di 30 gradi;
- una seconda analisi mirata all'individuazione di eventuali similitudini negli andamenti dei flussi veicolari suddivisi per fasce orarie;
- una ricerca delle regole maggiormente significative sia per ogni mese che per tutto il set di dati disponibili;
- la messa a punto del codice fuzzyTECH 5.54e utilizzando solo le regole precedentemente determinate.

Dalla prima fase è emerso che la frequenza di accadimento dei dati all'interno di ogni subset può considerarsi paragonabile, infatti, come mostra la Tab.3.3, l'ordine di grandezza rimane equiparabile. Tuttavia mentre nei range tra "0 e 60", "90-120", "150-180", "210-240", "270-300" e "330-360", la distribuzione degli stessi è pressoché identica, nelle restanti fasce luglio concentra i suoi valori nei range tra "60-90", "180-210" e "301-330" e giugno tra "120-150" e "240-270".
Anche la velocità del vento è del tutto similare nei gruppi corrispondenti; di fatto le eventuali differenze non potrebbero comportare da sole una diversa dispersione degli inquinanti ad eccezione della velocità massima registrata a giugno nel gruppo di direzione 0°-30° dove una differenza di più di 2 m/s rende possibile l'esistenza di disparate modalità di accumulo (vd Tab.3.4)
Da un più approfondito esame emergono però incoerenze tra le concentrazioni totali di monossido di carbonio dei due mesi: a parità di velocità, direzione del vento e classe di stabilità si verificano concentrazioni dell'inquinante minori per flussi veicolari maggiori (vd.Tab3.5).
Poiché tali situazioni si riferiscono a giorni feriali, del tutto paragonabili sia come traffico che come attività industriali, un simile andamento non trova spiegazione se non nella ricerca di particolari meccanismi che sono venuti ad instaurarsi nella complicata realtà urbana: esempi possono essere la sosta di veicoli davanti alla centralina che hanno costituito una barriera alla propagazione dell'inquinante oppure la loro fermata che al contrario hanno causato un accumulo dell'inquinante.

Non è stato comunque possibile definire quali dati costituissero difformità e tanto meno risalire alla causa, fondamentalmente per due motivi:

- i valori, se paragonati con l'andamento dello stesso mese da cui sono stati estratti, risultano coerenti;
- non esiste una memoria fotografica storica di come questi dati siano stati misurati

Limitando lo studio all'analisi dei range di variazione del monossido di carbonio rispetto alla direzione del vento, essi mostrano una totale coerenza tra i gruppi corrispondenti (vd.Tab.3.7)

Tab.3.3: frequenze di accadimento delle direzioni per intervalli di 30°

Direzione	Freq. Luglio *(Numerosità del set di dati)*	Freq. Giugno *(Numerosità del set di dati)*
0-30	3,86% (28)	4,30% (31)
31-60	5,10% (37)	4,85% (35)
61-90	12,67% (92)	8,74% (63)
91-120	13,36% (97)	13,18% (95)
121-150	17,22% (125)	20,39% (147)
151-180	7,85% (57)	7,77% (56)
181-210	2,62% (19)	0,97% (7)
211-240	7,30% (53)	7,21% (52)
241-270	15,29% (111)	18,03% (130)
271-300	11,16% (81)	11,65% (84)
301-330	2,75% (20)	1,66%(12)
331-360	0,82%(6)	1,25%(8)
	TOTALE: 100%(724)[3]	TOTALE: 100%(720)

Tab.3.4: Velocità minime e massime dei mesi di Luglio e di Giugno presenti in ogni subset

Direzione (gradi)	*Velocità min Luglio (m/s)*	*Velocità min Giugno(m/s)*	*| Differenza | (m/s)*	*Velocità max Luglio(m/s)*	*Velocità max Giugno (m/s)*	*| Differenza | (m/s)*
0-30	0,94	1,41	0,47	3,42	5,85	2,43
31-60	1,16	0,98	0,18	5,17	6,01	0,84
61-90	0,72	0,81	0,09	9,07	9,03	0,04
91-120	1,02	0,78	0,24	11,34	12,21	0,87
121-150	0,89	1,01	0,12	6,52	6,81	0,29
151-180	0,65	0,88	0,23	3,68	4,14	0,46
181-210	0,81	0,74	0,07	2,98	2,65	0,33
211-240	0,63	0,74	0,11	4,99	4,61	0,38
241-270	0,64	0,63	0,01	3,47	4,34	0,87
271-300	0,55	0,76	0,21	6,03	5,34	0,69
301-330	0,69	1,12	0,43	5,36	4,64	0,72
331-360	1,06	0,57	0,49	3,73	4,5	0,77

[3] La numerosità di luglio è inferiore a quella relativa a 31 giorni (744 rilevamenti) poiché 20 dati non sono stati resi disponibili dall'ARPAT

Tab.3.5: Esempi di incoerenze tra i dati(*)

Ora e data		direzione		velocità		cs		veic		COt	
G	L	G	L	G	L	G	L	G	L	G	L
h 9 25/06/02	h 9 2/07/02	139	138	1,33	0,89	b	b	880	1012	1,7	1,35
h 9 26/06/02	h 9 3/07/02	143	147	1,18	1,33	b	b	989	842	1,8	1,28
h 19 26/06/02	h 19 3/07/02	129	131	1,55	1,61	d	d	1251	1249	1,5	1,94
h 21 26/06/02	h 21 3/07/02	124	123	1,36	1,5	d	d	542	537	1,1	0,93
h 21 19/06/02	h 21 4/07/02	120	120	1,09	1,5	d	d	533	537	1,3	0,93
h 22 20/06/02	h 22 4/07/02	128	129	1,32	1,39	d	f	495	570	1,1	0,77
h 22 27/06/02	h 22 9/07/02	125	120	1,32	1,14	d	d	541	488	0,9	0,99

(*) E' stato omesso il contributo di monossido di carbonio dovuto alle attività antropiche presenti nell'area e al traffico della rete stradale escluso V.le Carducci poiché ritenuto non pertinente trattandosi di valori calcolati mediante il codice deterministico ISC3

Per quanto riguarda la classe di stabilità all'interno dei subset corrispondenti varia spesso la frequenza di accadimento e in alcuni casi anche il tipo (vd Tab.3.6). Nei gruppi maggiormente numerosi la distribuzione è però del tutto similare, le più rilevanti differenze emergono nei subset più poveri (vd Tab.3.3 e Tab.3.6)

Tab.3.6: Classi di stabilità e loro frequenza di accadimento nei mesi di giugno e di luglio

Direzione (gradi)	*Giugno* *Classi stab (frequenza%)*	*Luglio* *Classi stab (frequenza%)*
0-30	**b** (3,2) , **c** (22,6), **d** (54,8), **f** (19,4)	**a** (7,1), **b** (25), **c** (14,3), **d** (28,6), **e** (3,7), **f** (21,3)
31-60	**a** (11,4) ,**b** (37,2), **c** (34,3), **d** (11,4), **e** (5,7)	**a** (21,6), **b** (24,3), **c** (24,3), **d** (13,5), **f** (16,3)
61-90	**a** (43), **b** (25,4), **c** (1,6), **d** (15,8), **e** (1,6), **f** (12,6)	**a** (17,4), **b** (22,8), **c** (14,1), **d** (28,2), **e** (1,2), **f** (16,3)
91-120	**a** (28,5), **b** (15,8), **c** (8,4), **d** (33,7), **f** (13,6)	**a** (23,7), **b** (22,7), **c** (9,3), **d** (22,7), **e** (1), **f** (20,6)
121-150	**a** (27,2), **b** (23,8), **c** (9,5), **d** (24,5), **e** (0,7), **f** (14,3)	**a** (24), **b** (18,4), **c** (11,2), **d** (25,6), **e** (1,6), **f** (19,2)
151-180	**a** (8,9), **b** (14,3), **c** (14,3), **d** (55,4), **f** (7,1)	**a** (10,5), **b** (29,8), **c** (22,8), **d** (24,6), **e** (1,7), **f** (10,6)
181-210	**a** (14,3),**b** (14,3), **d** (57,1), **e** (14,3)	**a** (10,5), **b** (15,8), **c** (5,2), **d** (42,1), **f** (26,4)
211-240	**a** (2), **b** (15,4), **c** (4), **d** (21), **e** (9,6), **f** (48)	**a** (7,5), **b** (28,4), **c** (9,5), **d** (24,5), **e** (3,7), **f** (26,4)
241-270	**b** (6,9), **c** (0,7), **d** (26,4), **e** (25,3), **f** (40,7)	**a** (0,9), **b** (3,6), **d** (25,2), **e** (29,7), **f** (40,6)
271-300	**a** (1,2), **b** (13,1), **c** (13,1), **d** (25), **e** (10,7), **f** (36,9)	**a** (1,2), **b** (12,3), **c** (4,9), **d** (24,8), **e** (12,3), **f** (44,5)
301-330	**a** (8,3), **b** (25,1), **c** (8,3), **d** (8,3), **f** (50)	**b** (15), **d** (35), **e** (10), **f** (40)
331-360	**b** (12,5), **c** (12,5), **e** (12,5), **f** (62,5)	**d** (50) , **f** (50)

Tab.3.7: Range di variazione del COtot rispetto alla direzione del vento

Direzione (gradi)	*COtot Luglio*	*COtot Giugno*
0-30	0,35-1,70	0,3-2
31-60	0,52-1,70	0,3-2,1
61-90	0,26-1,94	0,2-2,1
91-120	0,18-1,99	0,1-2
121-150	0,23-2,39	0,1-2,4
151-180	0,27-2,07	0,8-2,5
181-210	0,40-2,05	1-2,5
211-240	0,21-2,07	0,2-2
241-270	0,19-2,24	0,2-2,4
271-300	0,17-2,23	0,2-3,0
301-330	0,17-1,75	0,2-1,5
331-360	0,48-1,32	0,2-1,2

Facendo invece un'analisi per fasce orarie (vd App.7), suddividendo le 24 ore nei gruppi "1-6,59", "7-10,59", "11-14,59", "15-16,59", "17-19,59", "20-22,59", "23-24,59" è risultato che l'andamento dei veicoli di luglio è simile a quello di luglio anticipato di circa tre ore per il primo gruppo, di due ore per tutti i restanti

3.4.1.1 Individuazione delle regole fuzzy maggiormente significative

La conclusione allo studio fatto in precedenza è che esiste una certa similitudine tra i dati analizzando separatamente le variabili e facendo riferimento a subsets definiti in base al range di variazione della direzione, ma ancora nulla può dirsi in merito alle regole fuzzy maggiormente significative per il contesto di analisi.
Tenendo presente quanto ottenuto al Cap.II, circa l'importanza relativa di ogni input rispetto all'output sono stati definiti il numero di ripartizioni del range di variazione di ogni variabile come riportato in Tab.3.8.

Tab.3.8: Ripartizioni delle variabili nel caso di utilizzo di 5 inputs e 1 output

	Variabile	Numero di ripartizioni	Valori
Inputs	Direzione del vento (°)	3 (low, medium, ,high)	low if 0<dir≤120 medium if 120<dir≤240 high if 240<dir≤360
	Velocità del vento (m/s)	3 (low, medium, high)	low if vel≤1 medium if 1<vel≤3 high if vel>3
	Classe di stabilità	3 (low,medium,high)	low if cs= a or b medium if cs= c or d high if cs=e or f
	Fondo (mg/m^3)	3 (low,medium,high)	low if fondo≤0,2 medium 0,2<fondo≤0,4 high if fondo>0,4
	Flusso veicolare (veicoli/h)	4 (low, mediumlow, mediumhigh, high)	low if veic≤100 mediumlow if 100<veic≤550 mediumhigh if550<veic≤800 high if veic> 800
Output	Monossido di carbonio totale (mg/m^3)	3 (low, medium, high)	low if COt≤0,6 mediumlow if 0,6<COt≤1,8 high if COt>1,8

Mediante tali ripartizioni possono essere individuate 972 regole fuzzy (vd.&2.3.3.3), non tutte con lo stesso grado d'importanza. Analizzando i vari casi occorrenti nei dati sono emerse circa 650 regole maggiormente significative. E' stato a questo punto allenato il codice fuzzyTECH5.54e implementando solamente le regole con grado di significatività >0; gli errori medi e massimi emersi in questa fase di training sono rimasti rispettivamente dell'ordine del 30% e del 70% e ciò ha reso non plausibile procedere nella fase di validazione.
E' stato ipotizzato che tali errori potessero dipendere da una ripartizione troppo "blanda" delle MFs degli inputs più importanti. Si è proceduto perciò ad una nuova configurazione delle ripartizioni dove nelle variabili esaminate non fosse presente il fondo, ritenendolo un parametro non vincolante. Questo ha permesso di ottenere quanto riportato in Tab.3.9.

Tab.3.9: Ripartizioni delle variabili nel caso di utilizzo di 4 inputs e 1 output

	Variabile	Numero di ripartizioni	Valori
Inputs	Direzione del vento	4 (low, mediumlow, mediumhigh,high)	low if 0<dir<90 mediumlow if 90<dir<180 mediumhigh if 180<dir<270 high if 270<dir<360
	Velocità del vento	4 (low, mediumlow, mediumhigh, high)	low if vel<1 mediumlow if 1<vel<2 mediumhigh if 2<vel<3 high if vel>3
	Classe di stabilità	3 (low,medium,high)	low if 1cs= a or b medium if cs= c or d high if cs=e or f
	Flusso veicolare	6 (verylow, low, mediumlow, mediumhigh, high, veryhigh)	verylow if veic<100 low if 100<veic<300 mediumlow if 300<veic<550 mediumhigh if 550<veic<800 high if 800<veic<1000 veryhigh if veic> 1000
Output	Monossido di carbonio totale	4 (low, mediumlow, mediumhigh, high)	low if COt<0,6 mediumlow if 0,6<COt<1,2 mediumhigh if 1,2<COt<1,8 high if vel>1,8

Ripetendo quanto fatto in precedenza sono emerse 312 regole maggiormente significative (vd.All.8), ma anche in questo caso il loro utilizzo non ha portato ad errori accettabili per le analisi da svolgere, anche se i loro valori sono diminuiti del 30%.
Anche in questo caso non si è ritenuto utile procedere nella fase di validation.

3.5 Conclusioni

Dallo studio svolto in questo capitolo non sono emersi criteri che potessero semplificare la trattazione dei dati in esame; rimane tuttavia confermata l'importanza della suddivisione dei range delle variabili attraverso la quale è possibile diminuire significativamente gli errori nella fase di training.

IV. NUOVO APPROCCIO MODELLISTICO

Lo studio svolto al Cap.III ha sostanzialmente evidenziato una distribuzione dei dati non uniformi. L'applicazione della cluster analysis e lo studio delle regole fuzzy maggiormente significative non hanno condotto a risultati accettabili per l'applicazione del modello a un caso reale poiché gli errori durante la fase di validazione del codice sono rimasti dell'ordine del 20%. E' risultato perciò necessario riesaminare il tutto partendo dalla ipotesi che i dati ricoprissero regioni di spazio ben definite, limitate ma non contigue. In queste gli intervalli di variazione delle variabili dovevano essere circoscritti in modo da consentire una maggiore precisione nella loro trattazione.

4.1 Ripartizione dei dati

Partendo da simili presupposti i dati sono stati suddivisi inizialmente in quattro gruppi tali che la direzione del vento fosse compresa tra 0° e 90°, 90° e 180°, 180° e 270°, 270° e 360° rispettivamente. Sono stati così ottenuti i seguenti raggruppamenti riportati nella Tab.4.1

Tab.4.1: Subsets ottenuti attraverso la ripartizione della direzione in quattro settori

Gruppo	Range di variazione della direzione	Numerosità
I	0≤dir≤90	283
II	90<dir≤180	577
III	180<dir≤270	372
IV	270<dir≤359	211

Poiché durante la fase di assessment è stato rilevato che risultati maggiormente significativi venivano ottenuti quando, per ogni variabile, il rapporto tra la deviazione standard e la media rimaneva inferiore al 40% (vd Cap.III), si è iniziato nell'imporre tale condizione al parametro più indicativo per la determinazione del monossido di carbonio. Costruendo la matrice di correlazione (vedi Tab.4.2), il parametro maggiormente correlato all'output è risultato, come era ovvio aspettarsi, il flusso veicolare.

Tab.4.2: Indici di correlazione tra le variabili

R	dir	cs	CO	vel	veic	COt
dir	-	-	-	-	-	-
cs	**0,37**	-	-	-	-	-
CO	**-0,01**	**0,00**	-	-	-	-
vel	**0,00**	**-0,07**	**0,001**	-	-	-
veic	**-0,43**	**-0,46**	**0,01**	**0,20**	-	-
COt	**-0,09**	**-0,2**	**-0,002**	**-0,4**	**+0,56**	-

Nei gruppi selezionati inizialmente attraverso la ripartizione della direzione del vento in quattro settori, il rapporto tra la deviazione standard e il valor medio del flusso veicolare rimaneva però sempre superiore al 40% (vd Tab 4.3) e ciò ha reso necessaria una ulteriore ripartizione ottenendo i subsets riportati nelle tabelle da 4.4 a 4.7

Tab.4.3: Rapporto della deviazione standard e della media del flusso veicolare nei gruppi individuati mediante la suddivisione della rosa dei venti in quattro settori

gruppo	Range della direzione del vento	numerosità	Deviazione standard (σ_{vei})	Media (μ_{veic})	σ_{veic}/μ_{veic} [4] (%)
I	0≤dir≤90	283	322	771	42
II	90<dir≤180	577	373	736	51
III	180<dir≤270	372	310	414	75
IV	270<dir≤359	211	329	449	73

Tab.4.4: Subsets con **0≤dir≤90** ottenuti imponendo una **σ_{veic}/μ_{veic}<30%**

0≤dir≤90				
Subset n°	Range di variazione dei veicoli	Numerosità subset	σ_{veic}	σ_{veic}/μ_{veic} [1] (%)
1	46-300	59	71	**38**
2	300-407	30	32,56	8,7
3	407-542	29	41	8,7
4	542-641	28	28,83	5
5	641-770	24	43,7	6,2
6	770-860	31	25,06	3
7	860-961	44	26,13	2,8
8	961-1061	58	30,9	3
9	1061-1161	27	26,44	2,5
10	>1161	34	38,17	3,4

Tab.4.5a: Subsets con **90<dir≤180** ottenuti imponendo una **σ_{veic}/μ_{veic}<30%**

90<dir≤180				
Subset n°	Range di variazione dei veicoli	Numerosità subset	σ_{veic}	σ_{veic}/μ_{veic} [5] (%)
1	48-144	**65**	32	**32**
2	144-249	29	31,7	16
3	249-380	32	29,61	11
4	380-460	31	25,9	6
5	460-501	31	12,06	2,5
6	501-535	30	10,87	2,1
7	535-581	39	13,9	2,5
8	581-655	27	24,09	4
9	655-716	27	19,03	2,7
10	716-753	28	11,09	1,5
11	753-806	30	18,36	2,3
12	806-851	32	13,49	1,6
13	851-886	32	11,84	1,3
14	886-917	36	10,5	1,16
15	917-933	27	6,18	0,7
16	933-953	26	5,63	0,6
17	953-980	33	8,96	0,9

[4] $\frac{\sigma_{veic}}{\mu_{veic}} \le 30\%$ dove $\sigma_{veic} = \sqrt{\frac{n\sum_{i=1}^{n} X_i - (\sum_{i=1}^{n} X_i)^2}{n^2}}$, $\mu_{veic} = \frac{\sum_{i=1}^{n} X_i}{n}$ e X_i = numero di veicoli nel gruppo i-esimo

[5] vd nota a piè di pag.46

Tab.4.5b: Subsets con **90<dir≤180** ottenuti imponendo una σ_{veic}/μ_{veic}**<30%**

90<dir≤180				
Subset n°	Range di variazione dei veicoli	Numerosità subset	σ_{veic}	σ_{veic}/μ_{veic}[6] (%)
18	980-1004	25	7,65	0,7
19	1004-1052	37	15,6	1,5
20	1052-1089	27	10,69	1
21	1089-1160	27	20,25	1,8
22	1160-1220	27	17,4	1,5
23	1220-1262	29	14,15	1,1
24	>1262	26	7,65	2,3

Tab.4.6 Subsets con **180<dir≤270** ottenuti imponendo una σ_{veic}/μ_{veic}**<30%**

180<dir≤270				
Subset n°	Range di variazione dei veicoli	Numerosità subset	σ_{veic}	σ_{veic}/μ_{veic}[2] (%)
1	39-57	26	5	10,5
2	57-86	29	8	12,6
3	86-135	33	15	13,2
4	135-175	31	14	8,7
5	175-207	31	10	5,2
6	207-272	27	21	8,9
7	272-340	29	21	6,7
8	340-439	27	36	9
9	439-517	28	23	4,8
10	517-594	28	18	3,2
11	594-726	29	42	6
12	726-803	29	22	2,8
13	803-951	32	43	4,7
14	>951	26	98	9,57

Tab.4.7 Subsets con **270<dir≤359** ottenuti imponendo una σ_{veic}/μ_{veic}**<30%**

270<dir≤359				
Subset n°	Range di variazione dei veicoli	Numerosità subset	σ_{veic}	σ_{veic}/μ_{veic}[7] (%)
1	46-83	27	8,8	14
2	83-151	27	18,9	15,2
3	151-213	27	17	9,7
4	213-342	26	39,7	14,9
5	342-450	30	35,6	8,9
6	450-650	31	67,3	12,3
7	650-893	34	73,8	9,5
8	>893	29	49,8	5

Dei 56 subsets, trentuno hanno mostrato avere $\sigma_{veic}/\mu_{veic} \leq 5\%$, quattordici hanno invece mostrato $5<\sigma_{veic}/\mu_{veic} \leq 10$, sette hanno evidenziato $10<\sigma_{veic}/\mu_{veic} \leq 15$ e solo quattro hanno avuto $\sigma_{veic}/\mu_{veic} > 15$

[6] vd nota a piè di pag.46
[7] vd nota a piè di pag.46

4.2 Training phases utilizzando i subsets

Per ogni subset è stato messo a punto un modello neuro-fuzzy impiegando l'80% dei dati per la fase di training del codice fuzzyTECH 5.54e ottenendo valori degli errori massimo e medio inferiori al 10 e all'1% rispettivamente (vd. Figg.4.1-4.2)

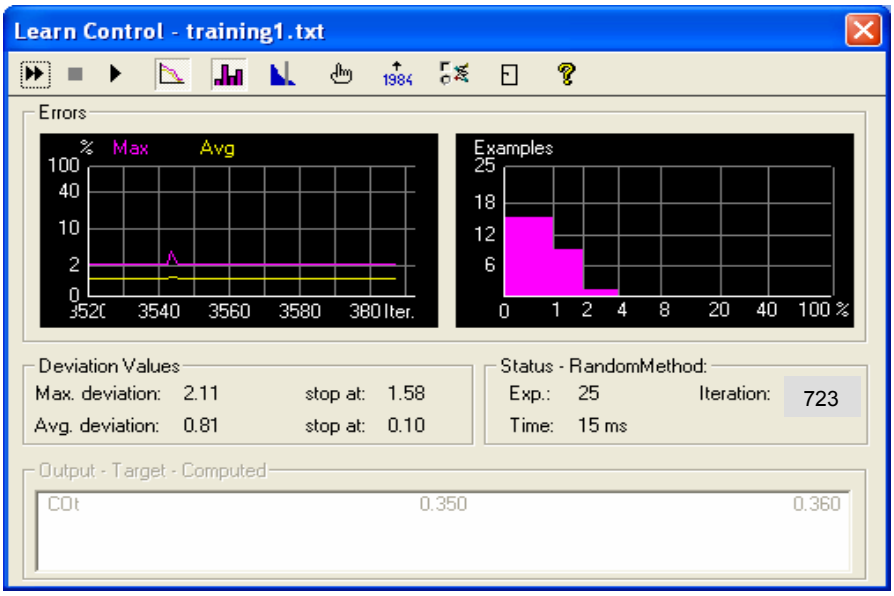

Fig.4.1: Esempio di training utilizzando un subset dove le "stop conditions" sono state ottenute per il raggiungimento della soglia massima di iterazioni preimpostata

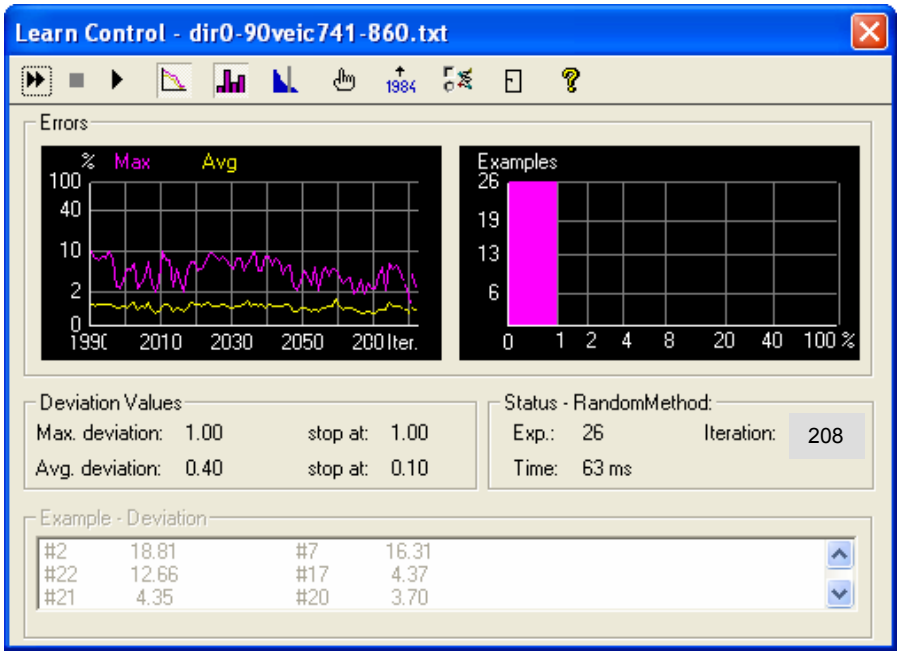

Fig.4.2: Esempio di training utilizzando un subset dove le "stop conditions" sono state ottenute per il raggiungimento della soglia di errore minimo calcolabile dal codice

Tali valori sono stati ottenuti alcune volte per il raggiungimento del numero di iterazioni massime preimpostato (1000), altre per il conseguimento della soglia di qualità propria del codice (errore massimo uguale a 1% o errore medio uguale allo 0,1%).

4.3 Validation phases

Per testare la validità dei risultati ottenuti durante la fase di allenamento sono stati utilizzati il 15% dei dati rimanenti dei sottogruppi, dati estratti in modo tale da essere rappresentativi di ogni subset esaminato. Circa il 90% dei subsets ha presentato un errore durante la fase di validazione minore del 18%; errori maggiori si sono verificati soprattutto nel gruppo di direzione del vento tra 0 e 90 (vd Tabs da 4.8 a 4.11)

Tab 4.8: Errori durante la fase di validazione per i subsets con **90<dir≤180**

Numerosità veicoli	Numerosità set training	σveic	σ_{veic}/μ_{veic} (%)	Numerosità set validation	$\varepsilon_{maxvalidation}$(%)
<144	56	32	32	10	2,84
144-249	26	31,7	16	4	8,49
250-380	27	29,61	11	5	8,6
381-460	26	25,9	6	5	7,73
461-501	26	12,06	2,5	5	4,58
502-535	**26**	**10,87**	**2,1**	**4**	**98,311**
536-581	34	13,9	2,5	6	5,49
582-655	24	24,09	4	4	1,22
656-716	24	19,03	2,7	4	1,23
717-753	24	11,09	1,5	4	1,86
754-806	26	18,36	2,3	4	1,67
807-851	28	13,49	1,6	5	7,15
852-886	28	11,84	1,3	5	6,56
887-917	31	10,5	1,16	6	5,5
918-933	24	6,18	0,7	4	2,58
934-953	23	5,63	0,6	4	1,06
954-980	28	8,96	0,9	5	3,82
981-1004	22	7,65	0,7	4	4,68
1005-1052	32	15,6	1,5	6	4,38
1053-1089	23	10,69	1	4	5,19
1090-1160	24	20,25	1,8	4	0,89
1161-1220	24	17,4	1,5	4	6,24
1221-1262	25	14,15	1,1	4	3,29
1263-1303	22	7,65	2,3	4	5,33

Tab.4.9: Errori durante la fase di validazione per i subsets con **270<dir≤360**

Numerosità veicoli	Numerosità set training	σveic	σ_{veic}/μ_{veic} (%)	Numerosità set validation	$\varepsilon_{maxvalidation}$(%)
<83	24	8,8	14	4	1,14
83-151	24	18,9	15,2	4	5,96
152-213	24	17	9,7	4	6,43
214-342	23	39,7	14,9	4	2,49
343-450	**26**	**35,6**	**8,9**	**5**	**57,6**
451-650	27	67,3	12,3	5	4,8
651-893	29	73,8	9,5	5	1,97
894-1027	26	49,8	5	5	7,46

Tab.4.10: Errori durante la fase di validazione per i subsets con **0≤dir≤90**

Numerosità veicoli	Numerosità set training	σ_{veic}	σ_{veic}/μ_{veic} (%)	Numerosità set validation	$\varepsilon_{maxvalidation}$(%)
<300	52	71	48	9	9
301-407	26	32,56	8,7	4	**129,9**
408-542	25	41	8,7	4	5
543-641	24	28,83	5	4	4
642-770	22	43,7	6,2	4	2
771-860	27	25,06	3	4	1,4
861-961	38	26,13	2,8	7	4
962-1061	52	30,9	3	9	2
1062-1161	24	26,44	2,5	4	**18,97**
1162-1311	30	38,17	3,4	5	**27,44**

Tab 4.11: Errori durante la fase di validazione per i subsets con **180<dir≤270**

Numerosità veicoli	Numerosità set training	σveic	σ_{veic}/μ_{veic} (%)	Numerosità set validation	$\varepsilon_{maxvalidation}$(%)
<57	23	5	10,5	4	1,75
57-86	26	8	12,6	4	2,8
87-135	28	15	13,2	5	7,32
136-175	27	14	8,7	5	2,75
176-207	27	10	5,2	5	5,47
208-272	24	21	8,9	4	3
273-340	26	21	6,7	5	5,6
341-439	24	36	9	4	6,87
440-517	**25**	**23**	**4,8**	**4**	**66,3**
518-594	25	18	3,2	4	1,6
595-726	25	42	6	4	5,86
727-803	25	22	2,8	4	6,5
804-951	27	43	4,7	5	4,66
952-1273	23	98	9,57	4	2,13

4.4 Analisi degli errori emersi nelle validation phases

I risultati ottenuti al §4.3 sono stati ulteriormente indagati graficando gli andamenti del monossido di carbonio calcolato rispetto ai valori reali di ogni gruppo in base alla ripartizione della. Tab.4.3. Le linee di tendenza hanno mostrato avere coefficienti angolari prossimi all'unità e coefficienti di traslazione dell'ordine del centesimo(vd Figg.4.4-4.7). In ogni gruppo i punti cerchiati in rosso(vd Figg.4.4-4.7), cioè quelli con errore > 18%, hanno distanze dalle rette di tendenza di almeno un ordine di grandezza maggiore rispetto agli altri. Ipotizzando una insufficiente conoscenza nelle regioni di spazio dove ricadono tali vettori è stata ripetuta la fase di allenamento inserendo nel gruppo di training gli stessi dati. Anche in questo caso però gli errori sono rimasti pressoché similari nonostante fossero valori trattati dal fuzzyTECH5.54e.
Riesaminando ogni subset (vd App.9) è emerso che i vettori con errore di validazione così elevato corrispondono, per quanto riguarda la direzione del vento e il numero di veicoli, ai limiti dei range delle variabili. Ciò ha fatto pensare ad eventuali problemi derivanti da border line.
Per confermare tale ipotesi sono stati dati gli inputs di tutti i vettori che appartengono ai limiti dei range delle variabili dette, vettori utilizzati già nella fase di allenamento, e gli errori trovati hanno sempre superato il 20% (vd.Tab.4.12) nonostante l'errore massimo indicato dal codice fosse risultato in ogni caso minore del 10%.

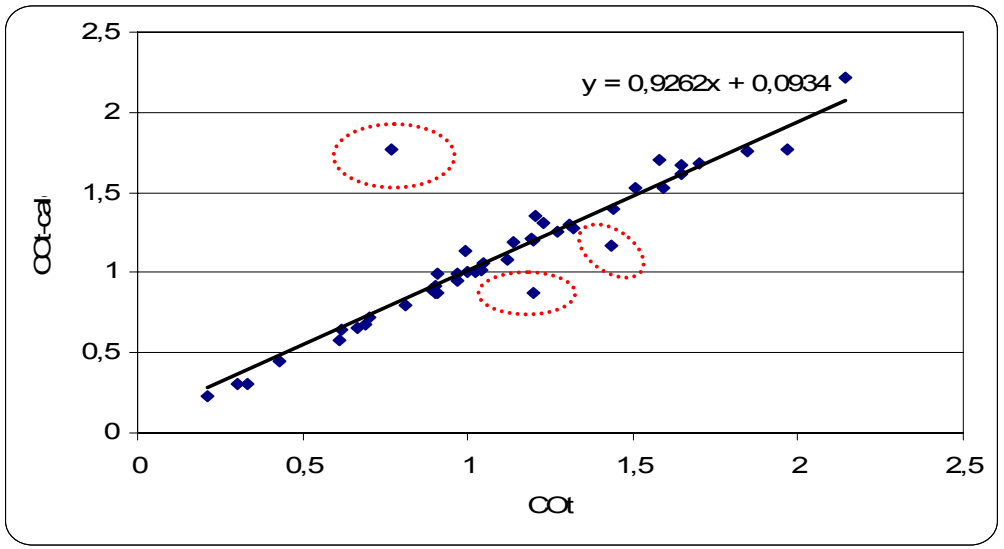

Fig.4.4: confronto tra dati calcolati e valori iniziali per il gruppo con 0≤dir≤90

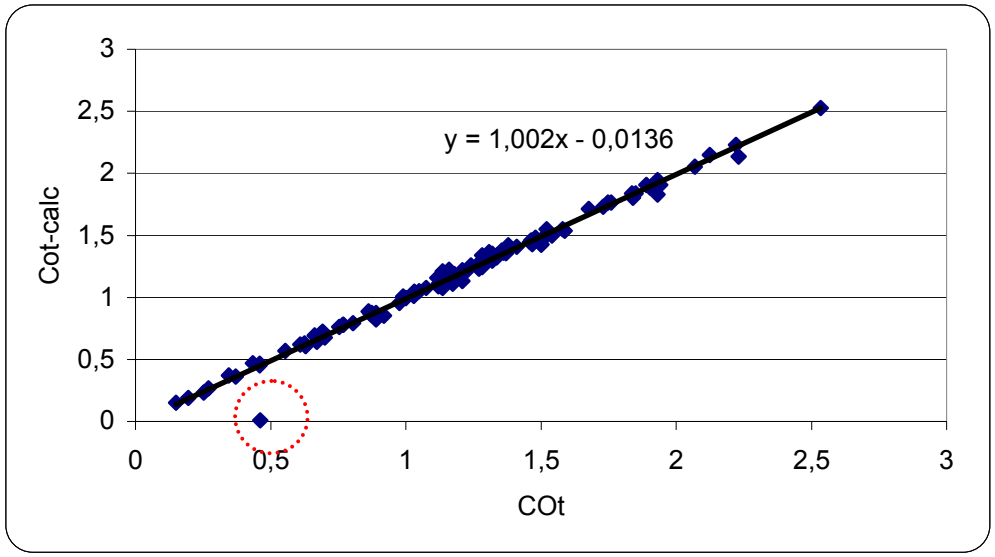

Fig.4.5: confronto tra dati calcolati e valori iniziali per il gruppo con 90<dir≤180

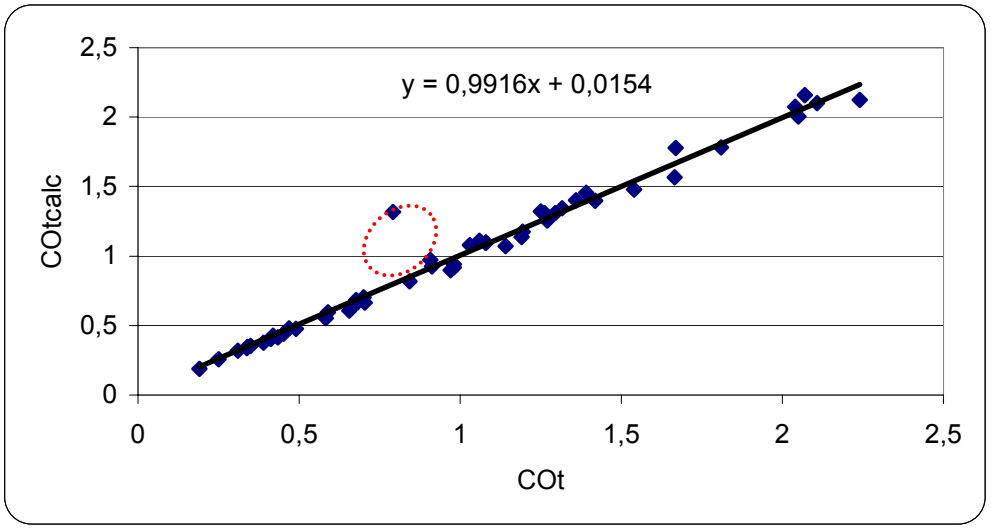

Fig.4.6: confronto tra dati calcolati e valori iniziali per il gruppo con 180<dir≤270

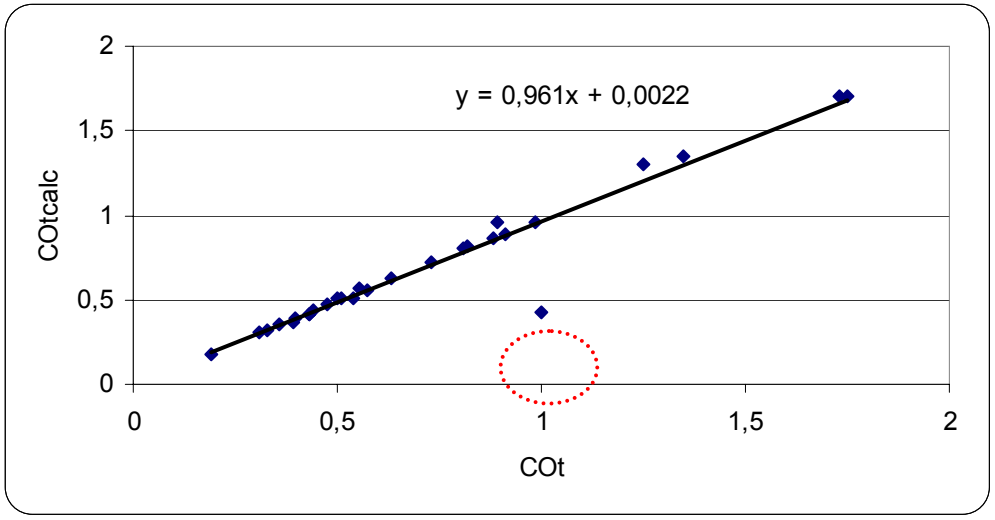

Fig.4.7: confronto tra dati calcolati e valori iniziali per il gruppo con 270<dir<360

Tab.4.12:Errori del codice fuzzyTECH5.54 sui valori limiti dei range della direzione del vento e del flusso veicolare

Gruppo	Subset	dir	cs	veic	fondo	vel	COt	COtcalc	$\varepsilon_{maxtraining}$(%)	$\varepsilon_{maxreale}$(%)
I (0<dir<90)	300<veic<407	**84**	1	**398**	0,11	2,3	0,47	0,93	2,95	**100**
	407<veic<542	**89**	4	**542**	0,25	1,73	0,89	1,42	5,78	**60,09019**
	641<veic<770	**83**	1	**761**	0,10	2,58	0,82	1,15345	6,3	**41,00856**
	1160<veic<1270	**86**	1	**1243**	0,10	2,13	1,21	1,54	2,1	**27,69486**
II (90<dir<180)	48<veic<144	**91**	6	**144**	0,09	6,73	0,54	0,1275	1,43	**76,34508**
	249<veic<380	**93**	6	**369**	0,28	2,43	1,02	1,2706	4,96	**24,93609**
	460<veic<501	**98**	6	**464**	0,35	1,62	0,77	0,5265	3,73	**31,80052**
	501<veic<535	**114**	4	**527**	0,06	7,6	0,46	0,0078	4,94	**98,31169**
	535<veic<571	**120**	1	**571**	0,11	2,34	0,56	0,7897	3,79	**40,01773**
	571<veic<655	**128**	2	**653**	0,11	2,45	0,76	0,50615	3,67	**33,31357**
	753<veic<806	**93**	1	**771**	0,15	1,8	0,75	1,35875	0,94	**82,13807**
	933<veic<953	**97**	1	**937**	0,10	2,47	0,91	1,24495	0,91	**36,80769**
	1160<veic<1220	**92**	1	**1199**	0,12	2,54	1,19	0,53825	3,45	**54,69276**
	1220<veic<1263	**101**	2	**1259**	0,18	1,68	1,35	1,89255	2,80	**40,39688**
	1263<veic<1303	**92**	3	**1296**	0,12	3,41	1,28	2,7493	3,83	**115,2937**
III (180<dir<270)	39<veic<57	**269**	6	**39**	0,04	0,88	0,28	0,430075	3,9	**53,59821**
	175<veic<207	**264**	5	**176**	0,07	1,19	0,66	1,1954	4,6	**82,22561**
	207<veic<272	**263**	4	**268**	0,06	1,01	0,59	1,03	6,15	**76,06838**
	340<veic<439	**258**	2	**341**	0,14	0,68	0,53	0,9	1,9	**69,17293**
	439<veic<517	**264**	5	**512**	0,30	0,82	0,79	1,31555	7,4	**66,31479**
	524<veic<726	**267**	4	**718**	0,04	1,69	1,18	0,5824	5,33	**50,64407**
	726<veic<803	**183**	2	**738**	0,12	0,87	1,31	2,07119	1,35	**57,62481**
IV (270<dir<360)	213<veic<342	**284**	4	**218**	0,02	1,11	1,39	0,7614	3,69	**45,30172**
	342<veic<450	**277**	5	**380**	0,15	0,87	1,00	0,42475	2,11	**57,60978**
	650<veic<893	**271**	4	**837**	0,01	4	0,89	1,5186	4,70	**71,01351**
	893<veic<1047	**276**	4	**1047**	0,12	1,54	2,66	1,3111	4,74	**50,76605**

Gli studi svolti durante i paragrafi precedenti hanno portato all'individuazione di un modello composito, costituito da 56 sottomodelli, ognuno dei quali dà risultati ottimali nei casi in cui la direzione del vento e il flusso veicolare non assumono i valori limiti dei loro range di variazione. L'uso dello stesso in fase prognostica ha dovuto però dapprima affrontare un altro aspetto riportato al §4.5: la localizzazione del nuovo dato rispetto al set utilizzato nella fase di training.

4.5 Modello composito

Come anticipato al § precedente, gli studi svolti hanno portato all'individuazione di un modello composito, costituito da 56 sottomodelli, ognuno dei quali dà risultati ottimali nei casi in cui la direzione del vento e il flusso veicolare non assumono i valori limiti dei loro range di variazione.
Volendo utilizzare il modello in fase prognostica, il problema emerso a questo punto è stato quello di stabilire se il vettore di input a disposizione costituisce o no un border line e come tale se i risultati ottenuti possono essere ritenuti affidabili. Mediante la direzione del vento e il numero di veicoli del nuovo dato (vd tab.4.13), è possibile individuare il sottomodello a cui far riferimento, ma niente ancora può dirsi circa i risultati forniti dallo stesso.

Tab.4.13: Individuazione del sottomodello da utilizzare mediante direzione del vento e numero di veicoli

Gruppo / Subset	Gruppo I (0<dir<90)	Gruppo II (90<dir<180)	Gruppo III (180<dir<270)	Gruppo IV (270<dir<360)
1	46<veic<300	48<veic<144	39<veic<57	46<veic<83
2	300<veic<407	144<veic<249	57<veic<86	83<veic<151
3	407<veic<542	249<veic<380	86<veic<135	151<veic<210
4	542<veic<641	380<veic<460	135<veic<175	210<veic<342
5	641<veic<770	460<veic<501	175<veic<207	342<veic<450
6	770<veic<860	501<veic<535	207<veic<272	450<veic<650
7	860<veic<961	535<veic<581	272<veic<340	650<veic<893
8	961<veic<1061	581<veic<655	340<veic<435	893<veic<1047
9	1061<veic<1161	655<veic<716	435<veic<517	-
10	1161<veic<1255	716<veic<753	517<veic<594	-
11	-	753<veic<806	594<veic<726	-
12	-	806<veic<851	726<veic<803	-
13	-	851<veic<886	803<veic<951	-
14	-	886<veic<917	951<veic<1281	-
15	-	917<veic<933	-	-
16	-	933<veic<953	-	-
17	-	953<veic<980	-	-
18	-	980<veic<1004	-	-
19	-	1004<veic<1052	-	-
20	-	1052<veic<1089	-	-
21	-	1089<veic<1160	-	-
22	-	1160<veic<1220	-	-
23	-	1220<veic<1263	-	-
24	-	1263<veic<1363	-	-

Per risolvere tale problema, avendo a disposizione dati stocastici multidimensionali, si è pensato di rappresentare ogni elemento del subset esaminato come un punto in un adeguato spazio metrico tale da rispettare nel miglior modo possibile le distanze originali /11/. Calcolando gli scarti dalla media dei vettori, essi rappresentano una matrice diagonale i cui autovalori costituiscono le componenti principali degli

elementi che meglio soddisfano quanto appena detto (vd App.10).
Imponendo un margine di errore (10%) tra lo scarto della matrice delle distanze reali D e quella delle distanze approssimate Δ, è possibile risalire a quali componenti possano essere ritenute sufficienti per una soddisfacente rappresentazione in un piano bidimensionale.

$$\mathrm{Max}\varepsilon_i = \max(\,|\,[D]-[\Delta(:;1:i)]\,|\,)$$

$$\mathrm{Max}\varepsilon \,/\, \mathrm{Max}\,[D] \leq 0{,}1$$

dove:

$\mathrm{Max}\varepsilon_i$: massimo errore commesso considerando le componenti da 1 a i del vettore esaminato;
[D]:matrice delle distanze reali dei punti esaminati;
[Δ(:;1:i)]: matrice delle distanze approssimate considerando i componenti;

$\mathrm{Max}\varepsilon / \mathrm{Max}\,[D]$ = percentuale del massimo errore ammissibile

Utilizzando tale metodologia, è così possibile identificare la posizione reciproca dei punti appartenenti al subset utilizzati per la fase di training e il nuovo vettore da impiegare in fase prognostica. Se il nuovo elemento appartiene alla stessa regione di spazio individuata dai dati impiegati nella fase di allenamento del sottomodello, quest'ultimo può essere ritenuto sufficientemente affidabile per l'applicazione in fase prognostica del nuovo dato. Il grado di accuratezza del risultato dipende:
- dall'errore commesso per la rappresentazione bidimensionale dei dati,
- dalla densità dei dati più prossimi a quello nuovo;
- dalla distanza che il nuovo dato ha rispetto a questi ultimi.

Si ritiene opportuno sottolineare che i fattori sopra elencati possono interagire in modo diverso a seconda del caso in esame. Non è possibile perciò a questo punto dello studio identificare un errore ideale di ogni singolo fattore che assicuri un grado di affidabilità dei risultati accettabile per l'utilizzo in fase prognostica del modello. Analisi storiche maggiormente rappresentative potrebbero fornire un valido supporto alla risoluzione del problema.

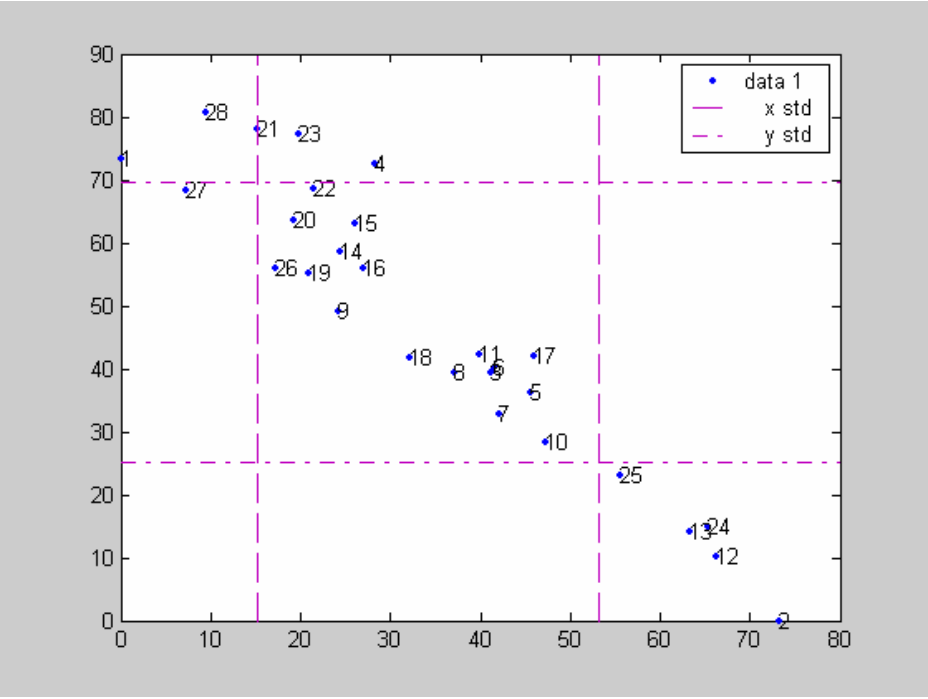

Fig.4.8: Esempio di rappresentazione dei dati di training e validazione di un sottomodello mediante la tecnica del Multidimensional Scaling.

4.6 Metodologia da applicare ad un sistema ibrido per una corretta analisi ambientale

A questo punto, sulla scorta degli studi e dei tentativi svolti in precedenza, possiamo arrivare a definire quello che, a nostro giudizio, sembra essere la migliore metodologia da usare per una corretta analisi ambientale.
Tale metodologia, riportata di seguito, può essere suddivisa in due parti fondamentali:

- una prima riguardante un approccio mirato alla valutazione dello stato dell'ambiente al momento dell'analisi
- una seconda rivolta alla previsione dell'inquinamento nel sito esaminato.

A) Analisi dell'inquinamento ambientale esistente
I passi da seguire sono:

a) Definizione dell'obiettivo (cosa studiare e perché, analisi costi-benefici,…)
b) Acquisizione dei dati
c) Verifica del teorema di Shannon
d) Data preprocessing (definizione degli outliers e selezione dei dati per la calibrazione off-line)
e) Scelta delle variabili d'ingresso (analisi dei fenomeni chimico-fisici, accorpamento di variabili mediante indici di correlazione o parametrici, contribution plots)
f) Analisi dei ranges di variazione delle variabili
g) Acquisizione delle regole fuzzy dalle esperienze degli esperti e search for structure in data
h) Clustering
i) Individuazione dei dati ottimali, consistente in:
    a. Analisi delle medie e delle deviazioni di ogni variabile
    b. Individuazione di sottogruppi nei quali la direzione ha un predefinito range (es. 0-90; 90-180; 180-270; 270-360)
    c. Individuazione della variabile maggiormente "correlata" all'output (matrice di correlazione tra inputs e output)
    d. Rielaborazione di ogni sottogruppo in ulteriori subsets che abbiano, rispetto alla variabile individuata al punto c), una σ/media<30% e una numerosità <40
    e. Utilizzazione, nel caso di subsets con σ/media>30% di sottogruppi con numerosità circa 50
j) Individuazione delle funzioni di appartenenza (MFs) tenendo conto che:
    a. l'uso di MFs complesse risulta spesso inutile e dispendioso; funzioni di appartenenza triangolari o trapezoidali rispondono in molti casi in modo soddisfacente alle esigenze del caso
    b. nell'effettuare la ripartizione delle MFs si deve tenere presente che:
        1. Variabili con deviazioni standard della popolazione paragonabili producono effetti simili sugli allenamenti
        2. Una maggiore ripartizione deve essere generalmente rivolta alle MFs delle variabili che mostrano maggiori deviazione standard della popolazione
k) Metodo random per l'estrazione dei dati;
l) Metodo random per l'affidamento dei pesi alle regole fuzzy;
m) Allenamento dinamico con soglia di massimo errore tra il 30 e l'1% e soglia di errore medio dell'0,1%;
n) Step With Dos, Step With Term e Winner pari a 0,1 , 1 e 1 rispettivamente
o) Limitare l'utilizzo dell' α-cuts solo se le regole superano le 200; in tal caso eliminare quelle che continuano ad avere il peso nullo dopo che il numero di iterazioni è pari a circa 5 volte la numerosità del set di dati

B) Previsione dell'inquinamento

I passi da seguire sono:

a) costruire la matrice delle distanze euclidee tra i vettori
b) costruire la matrice dei prodotti scalari degli scarti dalla media dei vettori
c) Ricavare la matrice degli autovalori
d) Costruire la nuova matrice delle distanze considerando i vettori rappresentati dagli autovalori del punto precedente

e) Rappresentare i vettori utilizzando il MDS verificando che il rapporto tra Maxerr calcolato tra due o tre colonne qualsiasi della nuova matrice delle distanze e il Max [D] ≤ 0,1
f) per la previsione utilizzare il modello che impiega i dati che risultano più vicini al nuovo vettore

4.7 Considerazioni finali

La metodologia definita al punto 4.6 presenta sufficiente elasticità nei limiti ragionevolmente accettabili per l'attività prognostica richiesta.
Infatti il modello consta di 56 sottomodelli la cui scelta viene determinata, durante la fase prognostica, dalla direzione del vento e dal numero di veicoli del dato da esaminare. Una volta individuato il sottomodello di riferimento, l'attendibilità dei risultati viene accertata in fase preliminare utilizzando la tecnica del "multidimensional scaling".
Con essa di fatto è possibile risalire ad una rappresentazione grafica dei dati multidimensionali e verificando la maggior o minor similitudine delle nuove circostanze rispetto a quelle già trattate dal modello durante la sua messa a punto, è possibile associare al valore trovato di CO un indice di "affidabilità", che risulterà tanto maggiore quanto le nuove circostanze saranno simili alle vecchie.
Conclusione dello studio svolto è senza dubbio che l'uso delle tecniche ibride si è mostrato essere un approccio corretto per la risoluzioni di problemi di analisi ambientali complesse.

Conclusioni

Il lavoro svolto durante questo dottorato ha portato alla messa a punto di una metodologia valida per l'analisi e la previsione dell'inquinamento dell'aria da traffico autoveicolare, in un contesto dove la realtà meteorologica associata ad una complicata urbanizzazione non rende possibile, se non con un notevole impiego di risorse umane e di calcolo, l'uso di codici deterministici classici.
La scelta dell'inquinante da indagare non limita la validità della metodologia, che è estendibile a qualsiasi sostanza di cui siano note le concentrazioni orarie ed offre perciò la possibilità di operare nell'ottica di un approccio integrato tra sviluppo e qualità di vita.
L'obiettivo è stato quello di trovare un buon compromesso tra precisione dei risultati e applicabilità a situazioni reali, dove spesso la trattazione dei dati non può avvenire in modo monodimensionale, ma mediante l'uso di vettori.
Nel caso specifico, ogni dato da utilizzare è risultato espresso da sei parametri (flusso veicolare del viale esaminato, velocità del vento, direzione del vento, classe di stabilità, fondo di CO dovuto alle attività antropiche presenti nell'area e al flusso veicolare di tutte le strade ad eccezione di quella studiata) e le analisi svolte hanno portato alla messa a punto di un modello composito. Il modello infatti consta di 56 sottomodelli la cui scelta viene determinata, durante la fase prognostica, dalla direzione del vento e dal numero di veicoli in circolazione sulla strada da esaminare. Una volta individuato il sottomodello di riferimento, l'attendibilità dei risultati viene accertata in fase preliminare utilizzando la tecnica del "multidimensional scaling".
Con essa di fatto è possibile risalire ad una rappresentazione grafica dei dati multidimensionali e verificando la maggior o minor similitudine delle nuove circostanze rispetto a quelle già trattate dal modello durante la sua messa a punto, è possibile associare al valore trovato della concentrazione di CO un indice di "affidabilità", che risulterà tanto maggiore quanto più le nuove circostanze saranno simili alle vecchie.
Conclusione dello studio svolto è senza dubbio che l'uso delle tecniche ibride si è mostrato essere un approccio corretto per la risoluzioni di problemi di analisi ambientali complesse. Purtroppo la carenza dei dati non ha permesso la costruzione di un modello interfacciabile direttamente con la realtà, per il quale la numerosità dei dati da trattare dovrebbe almeno ricoprire cinque periodi simili (es. giugno e luglio di cinque anni consecutivi).
Un possibile sviluppo della ricerca, per rendere tale modello un utile strumento per la pianificazione della circolazione urbana, potrebbe essere l'implementazione di dati più precisi relativi ai tassi di inquinamento del parco macchine circolante, al tipo di percorso stradale (presenza di semafori, rotonde viarie, parcheggi, aree di sosta, etc) e alle condizioni meteo su microscala, onde poter determinare i valori del flusso veicolare che assicurino il rispetto dei limiti imposti dalla normativa vigente.

Bibliografia

/1/ Kosko B., Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence, Prentice Hall, Englewood Cliffs, New Jersey, 1992.

/2/ Pedrycz W. and Card H.C., Linguistic Interpretation of Self Organizing Maps, In Proceedings of the IEEE International Conference on Fuzzy Systems, San Diego, pp. 371-378, 1992.

/3/ Nomura H., Hayashi I. and Wakami N., A Learning Method of Fuzzy Inference Systems by Descent Method, In Proceedings of the First IEEE International conference on Fuzzy Systems, San Diego, USA, pp. 203-210, 1992.

/4/ U.S. Environmental Protection Agency, User's Guide for the Industrial Source Complex (ISC3) Dispersion Model, Voll. I (User instructions) e II (Model Algorithms), EPA-454/b-95-0036

/5/ Ricerca Italiana- Università di Pisa, Roma, Palermo e Milano: Metodi di valutazione dell'incertezza nell'analisi di rischio di sistemi tecnologici complessi: applicazione ad una stazione di rifornimento idrogeno

/6/ E. Agostini, I. Ciucci, M. Mazzini, S. Strinati: Studio dell'inquinamento atmosferico da COV sul territorio di Livorno, con applicazione dei codici ISC3 e Caline 4-RL 1016 (03)

/7/. Santomauro L. (1977): Dinamica dell'inquinamento atmosferico da impianti industriali.

/8/ fuzzyTECH®5.5-User's Manual 2001

/9/ M.Veronesi, A.Visioli: Logica Fuzzy-Fondamenti teorici e applicazioni pratiche

/10/ H.C.Romesburg :Cluster analysis for researchers

/11/ W.Hardle, L.Simar: "Applied Multivariate Statistical Analysis"2003

/12/ A.M.Gadomski. TOGA: A Methodological and Conceptual Pattern for modeling of Abstract Intelligent Agent. Proceedings of the "First International Round-Table on Abstract Intelligent Agent". A.M. Gadomski (editor), 25-27 Jan.1993, Rome, Published by ENEA, Feb.1994.
As well as: http://wwwerg.casaccia.enea.it/ing/tispi/gadomski/gad-agen.html

/13/ A.M.Gadomski, IPK: Information, Preferences, Knowledge. 1997,

/14/ Simon Haykin, Neural Networks- a comprehensive foundations, second edition; Prentice-Hall, Inc.

/15/ Zadeh L.A., Fuzzy Sets, Information and Control, Vol. 8: pp. 338-353, 1965.

/16/ Zadeh L.A., Outline of a New Approach to the Analysis of Complex Systems and Decision Process, IEEE Transactions, System, Man, and Cybernetics, Vol.3, no.1, pp. 28-44, 1973.

/17/ Kosko B., Fuzzy Engineering, Upper Saddle River, NJ: Prentice Hall, 1997.

/18/ Cherkassky V., Fuzzy Inference Systems: A Critical Review, Computational Intelligence: Soft Computing and Fuzzy-Neuro Integration with Applications, Kayak O. et al (Eds.), Springer, pp.177-197, 1998.

/19/ Bayes T., An Essay Towards Solving a Problem in the Doctrine of Chances, Philosophical Transactions of the Royal Society of London, 53: pp. 370-418, 1763.

/20/ Dempster A.P., Upper and Lower Probabilities induced by a Multivalued Mapping, Annals of Mathematical Statistics, Vol. 38, pp. 325-339, 1967.

/21/ Shafer G., A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.

/22/ Judea P., Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers, USA, 1997.

/23/ Abraham A. and Nath B., Evolutionary Design of Neuro-Fuzzy Systems - A Generic Framework, In Proceedings of The 4-th Japan-Australia Joint Workshop on Intelligent and Evolutionary Systems, Namatame A. et al (Eds.), Japan, pp. 106-113, 2000.

/24/ Fukuda T. and Shibata M., Fuzzy-Neuro-GA Based Intelligent Robotics, In Zurada J.M. et al (Eds.), Computational Intelligence Imitating Life, IEEE Press, pp. 352-362, 1994.

/25/ Abraham A. and Nath B., Failure Prediction of Critical Electronic Systems in Power Plants Using Artificial Neural Networks, In Proceedings of First International Power and Energy Conference, CD-ROM Proceeding, Isreb M. (Ed.), ISBN 0732620945, Australia, 1999.

/26/ Sulzberger SM, Tschicholg-Gurman NN, Vestli SJ, FUN: Optimization of Fuzzy Rule Based Systems Using Neural Networks, In Proceedings of IEEE Conference on Neural Networks, San Francisco, pp 312-316, March 1993.

/27/ Lin C T & Lee C S G, Neural Network based Fuzzy Logic Control and Decision System, IEEE Transactions on Comput. (40(12): pp. 1320-1336, 1991.

/28/ Jang R, Neuro-Fuzzy Modeling: Architectures, Analyses and Applications, PhD Thesis, University of California, Berkeley, July 1992.

/29/ Nauck D, Kruse R, Neuro-Fuzzy Systems for Function Approximation, 4th International Workshop Fuzzy-Neuro Sy stems, 1997.

/30/ Tano S, Oyama T, Arnould T, Deep combination of Fuzzy Inference and Neural Network in Fuzzy Inference, Fuzzy Sets and Systems, 82(2) pp. 151-160, 1996.

/31/ Ajith Abraham: Intelligent Systems: Architectures and Perspectives, 2004

/32/ Abraham A., EvoNF: A Framework for Optimization of Fuzzy Inference Systems Using Neural Network Learning and Evolutionary Computation, 2002 IEEE International Symposium on Intelligent Control (ISIC'02), Canada, IEEE Press, 2002.

/33/ Abraham A., How Important is Meta-Learning in Evolutionary Fuzzy Systems?, In Proceedings of Sixth International Conference on Cognitive and Neural Systems, ICCNS 2002, Boston University Press, USA, 2002.

/34/ Abraham A. and Nath B., Evolutionary Design of Neuro-Fuzzy Systems - A Generic Framework, In Proceedings of The 4-th Japan-Australia Joint Workshop on Intelligent and Evolutionary Systems, Namatame A. et al (Eds.), Japan, pp. 106-113, 2000.

/35/ Rizzi, 1985; Mardia et al., 1979

/36/ Piccolo, 1984; Piccolo, 1986; De Thomasis et al., 2000